

動詞項構造シソーラスの構築

Construction of Thesaurus of Predicate-Argument Structure for Japanese Verbs

竹内 孔一*1

Koichi Takeuchi

岡山大学大学院 自然科学研究科

Graduate School of Natural Science and Technology Okayama University

This paper presents a thesaurus of predicate-argument structure for Japanese verbs to describe several levels of verb synonym groups. The proposed thesaurus defines verb concepts that can be shared among verbs with lexical decomposition description and analyzed example sentences. Since most of the verb senses correspond to word sense IDs in Lxceed then the thesaurus can be hopefully effective for verb sense disambiguation, which is a base technology that can bridge the gap between texts and the event ontologies organized based on descriptional logic.

1. はじめに

主に動詞を対象に言葉の言い換え関係を捉えられるための基本言語データとして、述語と項構造の関係をシソーラス形式で記事した辞書を人手で構築している*1。名詞の場合の類語と異なり、動詞を言い換える場合は動詞と係り関係にある句(項と呼ばれる)まで含めた関係を同定する必要がある。例えば

- 会社が 太郎を 社労士として 雇う/雇用する
- 会社が 太郎を 雇う/雇用する
- 会社が 社労士を 雇う/雇用する

では「雇う」と「雇用する」ではラゲは「太郎」など人を取る場合と、「社労士」など役割を取る場合があり、役割の場合には「として」という表現と言い換えが可能である。このような関係を扱う枠組みとして表層の助詞より一段階概念化した意味役割を仮定して、その組み合わせ動詞の概念と結びつける*2。

こうした述語項構造まわりの関係を記述しようとした言語資源は英語圏の言語学を中心に開発され、様々な枠組みが実データとともに提案されている(例えば EVCA [Levin 93], WordNet*3 (日本語では JWordNet*4), FrameNet*5 (日本語では JFN*6), LCS*7, VerbNet*8, など)。また日本語では EDR 電子化辞書*9 や IPAL [情報 86] などが構築されてきた。これに対して本辞書の記述枠組みの特徴は下記の通りである。

連絡先: 竹内孔一, 岡山大学大学院, 700-8530 岡山市北区津島中 3-1-1, 086-251-8178, 086-251-8178, koichi@cl.cs.okayama-u.ac.jp

*1 本研究は乾健太郎(東北大), 竹内奈央(言語アナリスト), 藤田篤(未来大)との共同研究の結果である。

*2 例えば 1 行目ならば「会社が [動作主] 太郎を [対象] 社労士として [役割] 」と記述する。この項に対する意味役割はいくつあるのかは解決しておらず様々な提案がなされていて [チャールズ, J. フィルモア 75][辻井 85][竹内 08][松林 10] 決着を見ていないが, 本研究では表現された見方を中心に言語処理を意識して付与を行う [竹内 08]。

*3 <http://wordnet.princeton.edu/>.

*4 <http://nlpwww.nict.go.jp/wn-ja/>.

*5 <http://framenet.icsi.berkeley.edu/>.

*6 <http://jfn.st.hc.keio.ac.jp/ja/>.

*7 <http://www.umiacs.umd.edu/~bonnie/>

LCS_Database_Documentation.html.

*8 <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.

*9 http://www2.nict.go.jp/r/r312/EDR/J_index.html.

- 動詞類義クラスに粒度を仮定して, 各粒度を類義語集合として階層的に取り扱う
- 類義クラスに項構造と事例を記述する
- 自他交替など動詞表現の機能的な異なりを記述する

こうした特徴により多面的に動詞間の関係を捉えることが可能になる。さらに実益的な側面として

- 自由に使えて Web 上で閲覧・利用できる

ことにより, どのような辞書の構造で, 現在どのようなインスタンスが登録されているのか, 更新された情報をすぐに確認できる仕組みにしている。

現段階で約 4400 語(約 7400 語義), 類義語集合(概念)の種類は 709 種類で, 分類は 5 階層, 必須の意味役割の種類は 71 種類である。最初に基本語意味データベース Lxceed*10 の語義を参考にしたため, ほぼすべての語義に Lxceed の語義 ID が付与されている。

本論文では, どのようなアイデアをシソーラスとして取り込み, その結果, どのような構造となったかについて説明する。

2. 動詞項構造シソーラスの構成

基本的なアイデアは, 複数の動詞を共通属性(概念)でまとめて, ざっくりどのようなタイプ(統語的, 意味的な振る舞い)があるのか俯瞰するとともに, 動詞の意味は個別であることから, そこから動詞間の差分構造が押さえられるのではないかとこのものである。こうした類義概念と差を記述することができれば, 操作的に動詞を選択したり, また文中にある動詞表現が他のどの表現とどう異なるのかがわかり, 文書理解に貢献できるのではと考えたためである。

例えば「本を買った」と「本を借りた」は本が手元にあるという結果状態では違いは無いが「買った」は所有権が移動しているが「借りた」は所有権は移動していない。このような関係を同時に記述するために, 上位の共通属性として【他者からの所有物の移動】*11 という概念を仮定して「買う」(「本を買

*10 http://www.gakken.jp/jiten/kihongo_db/.

*11 以後, シソーラスの分類カテゴリーである共通属性を概念とし【】で示す。なお実際の分類は階層分類であるが本文中では省略して記述する。

う」の語義での「買う」),「借りる」(「本を借りる」の語義の「借りる」)だけでなく,他者からの所有物の移動という属性を持つ他の動詞,例えば「レンタルする」「借金する」「借り入れる」をインスタンスとして所属させる。ここから下位概念として【獲得】と【借用】を仮定して,【獲得】には「買う」を所属させ,【借用】には「レンタルする」「借金する」「借り入れる」などの動詞を所属させる。ここですでに指摘しているように,動詞には複数の語義があるので,動詞そのもの(「買う」)を分類するのではなく,動詞のある語義に対する概念で分類する。よって語義の違いがわかるように,例文(「本を買う」)と項構造(「本/対象を買う」)をインスタンスとして書く概念に付与する(例えば「買う」の他の語義としては「けんかを買う」などある)*12。

以上がシソーラス構造を導入した基本的なアイデアである。これをより具体化するために言語学的な知見と言語処理での利便性を考慮してどのように動詞項構造シソーラスを構築したかについて,シソーラス構造,構造的意味記述,ならびに意味役割について記述する。

2.1 シソーラス構造

動詞間の共通する概念に粒度を仮定して,大きい概念からより細かな概念を構築する。この際,設計方針として下記を仮定した。

- (a) シソーラスの各ノードは概念であり,単語ではない
- (b) 概念間は上位下位関係のみに限定し,多重継承は行わない
- (c) ある語義が複数の概念に属する場合はインスタンスのみ
- (d) 階層の数は特に制限しない

まず(a)であるが単語がある単語の上位に来るということを経験的には仮定しない*13。「移動する」という動詞は「走る」や「スキップする」の上位としてまとめる方法も存在するかもしれないが,上位概念というのは下位概念を包括すべきと考えれば,単に「移動」ということを取り上げた動詞の意味が「スキップする」など特徴的な動作を包含するとは考えにくい。つまり,共通属性として【移動】を持つと考えれば「移動」,「走る」,「スキップする」という動詞を【移動】に分類する。

さらに,(b)でシソーラス内のノード(概念)間では多重継承を行わないようにノードを定義する。これは言語処理での利用を仮定した制約で,動詞(のある語義(インスタンス))は基本的にもっとも細分化されたノードに分類されるため,動詞のノードがわかれば,一意に概念がわかるようにするためである。一方で,動詞は多義であるので,(c)にあるように動詞は複数のノードに分類されることを許す。異なるノードに分類されることは概念が異なることを意味するので,動詞の語義の異なりを記述することができる。

上記の(a)(b)(c)の制約からシソーラスのノードは全て概念であり,上位下位関係のみで多重継承がないため,上位概念であるノードは必ず下位ノードに所属する全ての動詞を含むことになる。よって,ある動詞の語義がどのノードかわかれば,

*12 ここで,語義と概念の違いについて説明する。語義とは辞書などで見られる,各動詞が持つ異なる意味のことで,ここでは「買う」に対して「お金を払って手に入れる」と「けんかを仕掛ける」の2つを記述した。最低でも1動詞に1語義以上存在するため,語義数とは動詞の数より多くなる。それに対して,概念とは複数の動詞に共通する属性の分類であり,動詞の数より少ない。

*13 ただし,偶然そうなることは本シソーラスで起こる。この場合,上位語は単に上位階層のみに所属して,下位階層に現れなかった場合である。

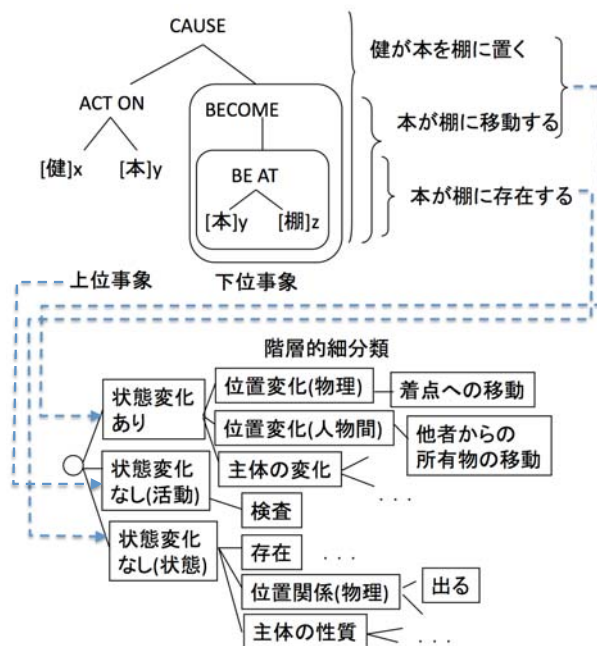


図 1: 語彙概念構造のシソーラス化による動詞概念の詳細化

上位概念は一意に決定できる。これにより,動詞の概念の曖昧性解消などを行えば,どのような概念属性があり,その動詞グループが存在するかわかる。この制約からシソーラスはノードの名前(概念名)に関係せず,ノードとインスタンス集合(動詞のある語義の集合)という関係はある共通属性を持つ類義語集合を多段的に提供できるため,概念名によらず単に動詞語義の類語集合を取り出すという使い方が可能になり,ユーザが細かな概念の設計を気にせず気軽に利用できる。

また,(d)の方針によって動詞を分類する概念をいくつでも細かく設定できるようにしている。基本的には動詞は個別の意味を持っているため,細かな粒度の概念が必要になったときに拡張できるようにしている。現段階では最大で5階層であるが必然ではない。

ここで図1にシソーラスの例を示す。シソーラスの第一階層には「状態変化あり」「状態変化なし(活動)」「状態変化なし(状態)」の3分類を仮定した。これは動詞分類における Vendler [Vendler 67]の分類を基にしており,動詞の基本分類と考えられる。この考えは語彙概念構造 [Jackendoff 90][影山 01]やモンタギュー文法に対する適応 [Dowty 79]が試みられており幅広く受け入れられている。全ての階層は紙面の都合上書けないので,ここでは省略する*14が例えば「本を買う」であれば,【状態変化あり-位置変化-位置変化(物理)(人間)-他者からの所有物の移動-購入】が省略しない場合の概念名である。つまり概念名は階層的に構成されており,このノードと同じ階層には【借用】(「借りる」),【奪回】(「取り返す」)などあり,所有物の移動のやり方についての違いを表している。また図2に示すように各ノードの上位階層にも動詞が存在し,下位の階層に進むにつれてインスタンスである動詞の数が少なくなり,概念が詳細化される。

*14 全てのデータは右のアドレスで配布している。http://cl.cs.okayama-u.ac.jp/rsc/data/.

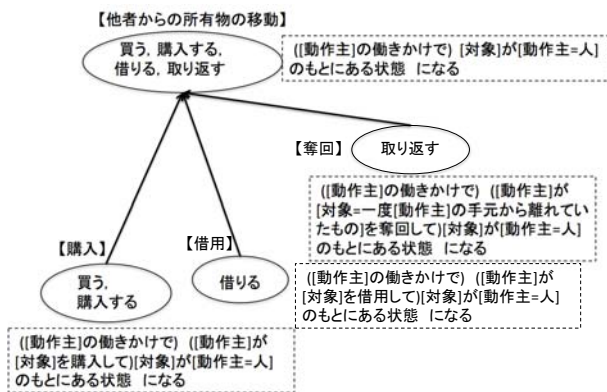


図 2: 上位と下位の階層と動詞および構造的意味記述の例

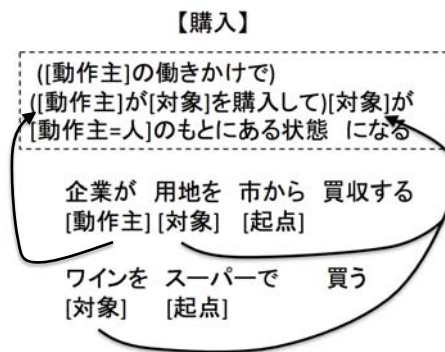


図 3: 1つの構造的意味記述(ノード)に対して複数の例文を対応

2.2 構造的意味記述(結果状態を中心に)

シソーラス内の各ノード(概念)の説明として項構造を詳細化した形で、意味の記述を行う。これは例えば WordNet のように各シソーラス内のノードの説明を文書で説明するのではなく、構造化して記述することで、(1) 構文内にあらわれる項同士の関係をゆるく形式化することで項の意味的な位置づけを行う、(2) 交替(alternation)や使役化といった生成的な異なりを記述することができる。まず(1)についてであるが、動詞間に共通する概念は動詞の項間の関係で記述できるというアイデアに基づいている。項とは表層のガワニではなく、一段階整理された動作概念の要素であり、それらの関係を動詞を利用して記述する。動詞を利用するとは「購入する」や「奪回する」などの動詞表現を構造内に取り込んで意味を記述することである。一方で語彙概念構造など言語理論では BECOME などの述語のみで記述する抽象化が行われてきたが、例えば「購入する」「奪回する」「借用する」に見られるように、BECOME として手に入れる変化を記述しても、手に入れる方法(お金を払ったか?)や形態(権利は移ったか?)、背景(もともと自分の物?)といった本質的な意味の異なりを記述するのは容易ではない。しかし、テキストを機械に理解させるにはこうした動詞の意味の違いを扱うことが必須であると考えられる。よって本提案手法では、具体的な動詞を利用して意味記述することで、より詳細な意味を表すことにした。また、こうしたデータからの整理を行うことで語彙概念構造など、理論に対する貢献ができることが期待される。

(2)について例をあげて説明する。図3の点線内に【購入】の構造的意味記述の例を示す。□内は項の意味役割を表しており、項間の意味的關係を構造的に定義することで【購入】の概念を表している。概念はおおむね、動作主、様態、結果状態という3つの構成要素から成り立っており、図3の点線内の各行に対応している。特徴的なのは「..ある状態」で【状態変化なし(状態)】を表し、「になる」で【状態変化あり】を表す。この構造化により、状態変化動詞と結果状態を結びつけることが可能になる。例えば図1の「本を棚に置く」「棚に移動する」という動作の結果状態と「棚に存在する」という状態は同じであり、こうした結果状態含意関係を記述している。

また、使役の脱着であるが、同様に「本を棚に置く」は他動詞による表現であるが、「本が棚に移動する」は自動詞による表現である。こうした自他の違いは概念の分類とは独立に操作的に扱えるように()で記述した、これにより、「本が棚の上に存在する状態」に至る自他の動詞を横断的に取り出すことが可

能になる。

2.3 意味役割をめぐる議論

意味役割とは項の種類を示しており、その役割としては、同じ項構造を持つ複数の動詞間で同様の意味を持つ項を同定することである。これを具体化するためには

- (1) 項としての役割: ある動詞の概念(シソーラスのノード)ではどういう項が出てくるか
- (2) 選択制限: 項が具体的な例文内でどうあらわれるかの対応

を記述する必要がある。ところが、項の種類が理論的に決着しておらず、格文法[チャールズ・J. フィルモア 75]や項構造[Grimshaw 90]など言語学的視点から数個程度が提案されたり、VerbNetのように付与と処理の観点から20数個程度が提案されたり、ガワニを中心に集約する方法[飯田 10]や書き分けることを中心に千種類を越える意味役割を用意する方法(FrameNet)が提案されている。また、処理の観点から千を超える意味役割は利用が難しく、集約したものを同時に利用した方がよいことが実験的に示されている[松林 10]。

本シソーラスでは、言語処理からの必要性和記述可能性の両方を視野に入れて整理することで意味役割に対する過度な期待を捨象することで意味役割を付与している。そこで過度な期待(上記「同様の意味を持つ項を同定」と)と、我々の方針について以下に整理する。

意味役割ラベルで期待されるのは、「統語的に異なっても現実世界において同一ならば同一の構造が付与されている」[辻井 85]ことであろう。文献[辻井 85]の例文を下記に引用する。

- その花が [Place] 密で [Obj] いっぱいになる
- 密が [Obj] 花に [Place] いっぱいになる

この2つの表現は【満たされる】という概念があるときに交替(alternation)が起こる典型的な構文である。しかしこうした「密が」と「密で」を同じラベルとして付与していく作業は実際の作業として困難であると筆者は考える。その理由として(a)手続きとして負担が高いこと、(b)人が文を認識する態度として見方が存在し、基本的にラベル程度で解決すべき問題ではないことを指摘する。

まず(a)であるが、結局言い換えの可能性を想像しながら意味役割を付与するという作業が必要である。上記の例はさほ

ど苦勞もないのは、意味役割の定義として上記(1)の概念の中の項の対応関係と言換えが対応しているためである。つまり【満たされる】という概念には【満たす場所 (Place)】と【内容物 (Obj)】が存在して、それが表層の助詞として違って現れても意味役割ラベルで吸収できる。ところが、この概念内の位置づけと、意味的に同じというのは必ずしもうまく一致しない。例えば

- 石原氏 [経験者] が都知事に [役割] 再選する
- 石原氏 [動作主] が都知事に [役割] 就任する

では「石原氏」は動作主体か(つまり [動作主]), 主体だがコントロール不可か(つまり [経験者](他には「彼 [経験者] は風邪を引いた」))の違いが上下の文で異なっているが [役割] に着任する主体であることに代わり無い。これは syntactic な見方の違いから意味役割のラベルが異なってしまうのが原因であるが、こうした違いを意味役割ラベルで吸収しようとする、他の動詞の言換えを常に考慮して、ラベルを定義しなおす必要があり、付与が収束しない。

また同時に (b) の説明にもなるが、syntactic な見方と「着任する主体」という意味的な見方は別の見方であり、こうしたものをラベルで統一できる保証がないと考えられる。もっとわかりやすい例では「売る」「買う」がある [Fillmore 82][Takeuchi 10]。「太郎が古着を次郎から買う」と「次郎が古着を太郎に売る」は起こった事象は同じで、見方の異なる表現(「太郎」主体か「次郎」を主体として考えるか)であるが見方の異なりはある事態に対する取り上げ方の異なりで、ラベルのみで解決できる問題ではない。

こうした問題点を受けて、本シソーラスでは実世界に対するマップを意味役割で行うのではなく、実世界で起こった実体を推論するための基本データとして、意味役割を付与する。つまり、ある概念の見方で意味役割ラベルを付与し、言換えなど他の言換えとの関連は alternation の範囲までする。

概念と意味役割との関係であるが、意味役割ラベルは事例を集約したのちに再考するものと考え、概念の必須の意味を上記の構造的意味記述に取り込み、それ以外は、事例に対して付与を行う。よって図3にあるように、1つのノード(概念)には必須と思われる少数の意味役割を利用して概念を記述し、それ以外に出現する意味役割は例文とセットで記述し、必須でない物がどの程度現れるかを示している。この事例とのセットにより、上記(2)の選択制限に対する情報を提供する。現在意味役割の種類は71種類あるが、フラットでなくおおまかに分類が存在し、まだ整理中ではあるが半分以下の上位分類にまとめられると考えられる。

3. オントロジーとの関係

近年、概念を定義して概念間の意味関係を論理的に整理することで知識を記述するオントロジーの研究が進み、具体的なイベントオントロジー [兼岩 10] が日本語の動詞に対して定義されてきている。オントロジーと本シソーラスとの違いはオントロジーが概念間の正確な記述的操作を含む定義に重きを置くのに対して我々はテキストから概念に変換する曖昧性解消に重点をおいている点である [Tsujii 05]。つまり役割が異なっており、提案するシソーラスによりテキストから概念へのマップ(語義曖昧性解消)を高い精度で行うことができれば、イベントオントロジーで定義されている記述論理を利用する事が出来るため、より深い意味処理を行うことが期待できる。

4. 今後の展望

現在もシソーラス内の動詞(述語)要素の追加と事例の追加を行っている。また言語処理でシソーラスの利用することでシソーラスに対する評価を行う予定である。

参考文献

- [Dowty 79] Dowty, R. D.: *Word Meaning and Montague Grammar*, Dordrecht: Reidel (1979).
- [Fillmore 82] Fillmore, : *Frame Semantics*, pp. 111-138, Hanshin Publishing Corporation (1982).
- [Grimshaw 90] Grimshaw, J.: *Argument Structure*, MIT Press (1990).
- [Jackendoff 90] Jackendoff, R.: *Semantic Structures*, MIT Press (1990).
- [Levin 93] Levin, B.: *English Verb Classes and Alternations*, University of Chicago Press (1993)
- [Takeuchi 10] Takeuchi, K., Inui, K., Takeuchi, N., and Fujita, A.: A Thesaurus of Predicate-Argument Structure for Japanese Verbs to Deal with Granularity of Verb Meanings, in *The 8th Workshop on Asian Language Resources*, pp. 1-8 (2010).
- [Tsujii 05] Tsujii, J. and Ananiadou, S.: Thesaurus or Logical Ontology, Which One Do We Need for Text Mining?, *Journal of Language Resources and Evaluation*, Vol. 39, No. 1, pp. 77-90 (2005).
- [Vendler 67] Vendler, Z.: *Linguistics in Philosophy*, Cornell University Press (1967).
- [チャールズ J. フィルモア 75] チャールズ J. フィルモア (田中春美, 船城道雄訳): *格文法の原理*, 三省堂 (1975).
- [影山 01] 影山 太郎: *動詞の意味と構文*, 大修館書店 (2001).
- [兼岩 10] 兼岩 憲, 岩爪 道昭: セマンティック Web のためのイベントオントロジー, *コンピュータソフトウェア*, Vol. 27, No. 5, pp. 1-13 (2010).
- [松林 10] 松林 優一郎, 岡崎 直観, 辻井 潤一: 自動意味役割付与における意味役割の汎化, *自然言語処理*, Vol. 17, No. 4, pp. 59-90 (2010).
- [情報 86] 情報処理振興事業協会技術センター情報処理振興事業協会技術センター: *計算機用日本語動詞辞書 IPAL(辞書編)* (1986).
- [竹内 08] 竹内 孔一, 小山 照夫: 動詞の語義と意味役割を付与したタグ付コーパスの作成, *電子情報通信学会言語理解とコミュニケーション研究会 NLC2008-77*, pp. 19-22 (2008).
- [辻井 85] 辻井 潤一, 山梨 正明: 格とその認定基準, *情報処理学会自然言語処理研究会*, No.48, pp. 1-7 (1985).
- [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).