

多義語の語義分布と語義間の派生関係調査の試み—相の類を中心に

加藤祥¹・田邊絢²・浅原正幸¹・古宮嘉那子²・新納浩幸²

1 国立国語研究所, 2 茨城大学

1. はじめに

本発表は、意味情報を付与したコーパスを用い、多義語とされる語（相の類の144語、386義）の意味分布の実態を調査した結果を報告する。まず、語義の頻度調査を行い、先行研究によって内省的に示されてきた代表義や基本義との関係を確認した。次に、用例から意味の派生関係を分析し、派生関係情報を付与する作業を行った。本発表では、調査の結果、各語義における高頻度語義を取得し、先行研究との関係を示すとともに、用例に基づき各義の派生関係を分析した例を示す。本調査により、高頻度語義と派生関係の傾向が明らかとなったほか、語のプロトタイプの意味の認定において考慮すべき問題が確認された。今後、本結果を基礎データとし、多義語の用例を用いた各義の派生関係判定実験を行う予定である。

2. 関連研究と本研究

多義語分析の課題として、靱山(2001,2002,2003)は、①プロトタイプの意味の認定、②複数の意味の相互関係の明示、③複数の意味すべてを統括するモデル・枠組みの解明の3つを挙げる。すなわち、複数の意味のなかに、「最も基本的であり、慣習化の程度・認知的際立ちが高いといった」特徴を備えた意味を認定すること、複数の意味の関連の実態を分析すること、多義構造全体における個々の意味の位置づけを明示することなどが求められると考えられる。

多義語の語義は当該言語において定着した意味(Cruise, 2004)であり、一般に辞書へ記載されるものと考えられる。しかし、複数ある語義のうちいずれの語義がプロトタイプの意味(代表的、あるいは基本的など)なのか、という点については、用法上の制約による判定や、多義性のネットワークからの認定、心理学的判定など(靱山, 2003)様々な手法がとられるが、判断に差が生じるものもある。語義の掲載順を見ても、出現年代順(例: 古い順『広辞苑』, 現在通用している順『大辞林』など)や書き手による重要度判断順(例: 意味の関連『ジーニアス』第3版まで)、頻度順(『COBUILD』, 『WISDOM』など)と、辞書によ

って異なる。

また、多義語の代表義や基本義は、例語を挙げた分析が数多くなされてきたが、網羅的に認定を行ったデータとして、『分類語彙表』(国立国語研究所, 1964)を用いた研究がある。国立国語研究所(1984)の『日本語教育のための基本語彙調査』は、日本語シソーラス『分類語彙表』を用い、語義の観点から日本語教育・言語教育の専門家によって基本語彙を選定したものである。すなわち、『分類語彙表』項目に基づき、意味上適切とされる基本語が選定されているため、収集された語は基本的な意味において選定されていると考えられる。さらに、山崎・柏野(2017)は、『分類語彙表』増補改訂版(国立国語研究所, 2004)の多義語に人手で代表義を付与する作業を行った。よって、『分類語彙表』の項目となっている多義語のすべて(14,102語; 山崎・柏野(前掲))について、代表義と判定された語義の情報が付与されている。但し、これらの基本義、代表義については、複数人の判断による、あるいは恣意的にならないように留意するための作業基準が設けられているものの、専門性を有した人の内省的判定であり、高頻度の語義であるのか、あるいは一般的にプロトタイプののであるのかなどの検証結果がほしい。

そして、同音語との区別として、多義語の語義はいずれも何らかの関連を有していると認識されるものである。しかし、基本義と派生義の関連性は、シネクドキなどによる規則的または体系的な場合と、メタファーをはじめとする転用によって不規則的あるいは非体系的な場合がある。語義間の派生関係を個別的に考察した研究は多いが、網羅的な調査や実態的な傾向分析の実施はこれまで困難であった。さらには、複数の意味を統括するモデル(イメージスキーマに基づくモデル(Lakoff, 1987など)、拡張とスキーマに基づくモデル(Langacker, 1987など)など)が考案されてきたが、大規模に検証することも難しい。

そこで、本発表では、コーパスにおいて多義的な振る舞いの見られる語について、各語義の頻度を調査し、先行研究に示される代表義や基本義との関係を明らかにするとともに、多義語の各語義

に派生関係情報を付与した結果を報告する。今後、語義間の派生関係に基づき、複数の意味を統括するモデルを実証的に検討したい。

3. 調査

3.1 データ

『分類語彙表』の分類項目¹（語義情報）を付与したコーパスとして、加藤ら（2018）の『現代日本語書き言葉均衡コーパス』（以降BCCWJ）の新聞・雑誌・書籍の各サブコーパス（約35万語）に対するアノテーションデータがある。この分類語彙表番号付与済みコーパスでは、多義的な語についても、文脈上適切な意味分類が一意に付与されている。よって、本コーパスを用い、多義語の各語義の頻度を調査することが可能である。

また、本発表では、『分類語彙表』の相の類の語について、実際にコーパス上で多義的に使い分けがなされている100語程度の語を調査対象語として選定した。多義語として多義的な意味が認定されている語（1,246語）でも、実際には用例が一意の場合もあった。たとえば、形容詞「青い」は、『分類語彙表』において3つの分類項目（3.3421（精神および行為-行為-才能）、3.5020（自然現象-自然-色）、3.5701（自然現象-生命-生））に分類されているが、本調査²において取得された用例は「3.5020（自然現象-自然-色）」のみである。反対に、『分類語彙表』では多義語として認定されていないが、実態として新語義が付与されている場合もある。たとえば、形容詞「偉い」は、『分類語彙表』において単義の扱い（3.3422（精神および行為-行為-威厳・行儀・品行）のみ）であるが、分

¹ 『分類語彙表』を構成する分類階層の1つで、895の分類項目があり、それぞれ以下のような5桁の数字で表される。

補表 分類番号の構造

（例：この（分類番号：3.1010））

類	部門	中項目	分類項目
用 (3)	関係 (.1)	真偽 (.10)	こそあど (.1010)

² なお、本調査に用いたデータ外には、「尻が青い」の慣用句として用いられる「青い」用例がBCCWJ 3,486例中5例あるほか、「まだかすかに青さの残る恋を甘酸っぱく感傷的にした。（サンプルID：PB29_00214）」という3.5701にあたりと考えられる用例は存在する。また、「青いトマトを平気でもいだりする（サンプルID：PB3400067）」のような判断が揺れる可能性のある用例も見つかっている。

類語彙表番号付与済みコーパスでは、多義語の振る舞いが確認される（3.1920（抽象的關係-量-程度）が付与された語もある）。そこで、相の類の語のうち、コーパス上で2種類以上の語義使用が確認され、各語義について用例の頻度が5以上取得可能な語を抽出したほか、多義用例の取得が可能な高頻度語の計144語、386義を調査対象とした。

3.2 作業

①意味分類別に用例数を取得し、頻度を手掛かりとして代表義の判定を行った。山崎・柏野（2017）の認定する代表義との対照を行う。また、『日本国語大辞典³』（出現年代順：古い→新しい）を確認し、②多義のうち最も古くから認められる意味をマークし、基本義の補助情報とする。なお、基本義の判定には、『日本語教育のための基本語彙調査』で基本語とされている語義も参照した。

さらに、③各語義について、いずれかの他義から派生関係が認められる場合には、作業者1名（第2著者）が、元となったと考えられる語義と、派生関係について情報を付与した。派生関係は、概念領域が同じ場合は規則的な派生と考え、メトニミー・シネクドキ的な情報⁴を付与し、概念領域が同じではない感覚がある場合は不規則な派生と考え、メタファー（「転用」）と付与した。

3.3 結果

結果の一部を表1に示す。先行研究で代表義と認定されていた語義は、必ずしも用例数において最多ではない（例：表1の「若い」や「遅い」）。また、出現例が古い（『日本国語大辞典』において上位語義として記載がある）語義が、必ずしも代表義と認定されているのではないことも明らかとなった。また、多義語によって、派生関係の種類も様々な傾向が確認された。

たとえば、「柔軟」の多義において、『日本国語大辞典』における最も古い語義は3.5060（自然現象-自然-材質）であり、代表義（山崎・柏野，2017）であるとも判断されているが、用例数を見ると、3.1500（作用-作用・変化）が最も高い頻度となっ

³ <https://japanknowledge.com/library/>

⁴ 全体（上位カテゴリ）→構成要素、部分（下位カテゴリ）→全体、代替表現、具体→抽象など、みとめられた関係を付与した。メトニミー・シネクドキ・メタファーの区別には諸説あるため、本調査では比喩的な分類は特定していない。

ている。以下の(1)は「柔軟」に3.5060の付与された例、(2)は「柔軟」に3.1500が付与された例である。材質の意味の用例が取得される文脈は限られ、抽象的な内容の文脈において作用・変化の意味で用いられる例が多いためである。なお、比喩的な意味である3.3680(精神および行為-待遇-待遇・礼など)が付与された(3)のような例もある。

(1) ウロルーメ(写真1-3)は、クロムニッケル合金製メッシュ状ステントで、柔軟性があり自己拡張します。(サンプルID:PM26_00003, 下線は著者による; 以下同様)

(2) また、糖尿病による性機能障害の患者にバイアグラを処方することも、日本と似た保険制度を持つフランスでは、一定の制限をかけながら柔軟に対応しているのに、日本では現在、禁止されている。(サンプルID:PN1c_00006)

(3) 中国が東南アジア諸国との自由貿易交渉で、農産物取引においても柔軟な姿勢を見せているのに、日本が2国間交渉で、農業改革に本腰を示さないとすると、(サンプルID:PN3b_00007)

これらの用例から、3.5060(自然現象-自然材質, 例文(1))という具体的な語義から抽象的な語義3.3680(精神および行為-待遇-待遇・礼など, 例文(2))が派生しており、また、3.5060(自然現象-自然材質, 例文(1))からの比喩的な転用として、3.3680(精神および行為-待遇-待遇・礼など, 例文(3))が派生したという関係が考えられる。さらに、用例の意味的な分布において、3.3680(精神および行為-待遇-待遇・礼など)という抽象的な語義が最も高頻度となっていることが確かめられた。

同様に、「遅い」においても古くかつ代表義と判断されていた語義3.1660(抽象的關係-時間-新旧・遅速, 例文(4))から3.1913(抽象的關係-量-速度, 例文(5))が派生していると考えられ、より抽象的な語義が高頻度で用いられるという現象が見られる。

(4) パリでは芝居もそうだが、こういう催し物は夜の九時開場というのが普通で、日本にくらべると大変に遅い。(サンプルID:PB47_00014)

(5) 2. 5Gは、現行のネットワークに機能を追加してパケット通信を可能にしたもので、データ伝送速度は3Gより遅い。(サンプルID:PM25_00067)

また、「若い」では、3.5701(自然現象-生命-生, 例文(6))が代表義と判断されている語義であるが、3.1660(抽象的關係-時間-新旧・遅速, 例文(7))が古く、用いられる頻度も高くなっている。時間的な3.1660(抽象的關係-時間-新旧・遅速)から年齢等に関係のない3.5701(自然現象-生命-生)に派生したものと考えられる。なお、部分から全体を言うという派生関係と考えたものの、関係性の判定には疑問が残っている。判定が困難な語では、分類語彙表番号付与の段階からも、揺れの生じていることがわかっている。(8)は3.1660が付与された例である。このような例では、意味の判定が困難であるため、3.1660と判断された用例と3.5701と判断された用例が生じている。今後、用例の意味判定実験を行い、分類語彙表番号付与済みコーパスにおける語義の精度を高めることを予定している。

(6) もしあなたがまだ若かったら、人生の荒波にもまれた経験など少ないでしょうから、そういう

表1 作業結果例

語彙素	分類語彙表番号	用例数	日国上位語義	●:用例数最多 △:代表義判定	派生関係
柔軟	3.1500	6		●	【3.5060から具体→抽象】
柔軟	3.3680	1			【3.5060から転用】
柔軟	3.5060	1	◎	△	
遅い	3.1660	8		●	【3.1913から具体→抽象】
遅い	3.1913	4	◎	△	
若い	3.1660	42	◎	●	
若い	3.5701	17		△	【3.1660から部分→全体?】
更に	3.1612	23	◎		
更に	3.1920	65		●	【3.1612から部分→全体】

境地は理解できないでしょう。(サンプル ID : PB11_00013)

(7) 広くはないが若い娘なら喜びそうな部屋で、可愛いカバーのかかったベッドが二つ、クラシックな形の椅子が二つ置いてある。(サンプル ID : PM31_00020)

(8) 若くて元気があれば夢を見ることも、それを実現することも確かにできる。(サンプル ID : PM12_00014)

このほか、表1で示した「更に」は、代表義判定と高頻度語義が一致しているが、古い語義ではなかった例である。

このように、本調査の結果において、古い語義、代表義、高頻度語義のいずれかが一致していない語の例が多く見られた。むしろ、古い語義と代表義、高頻度語義が一致した例は、本調査の対象語144語中50語のみであった。いずれをプロトタイプの意味ととらえるべきであるのか考慮する必要がある。

4. まとめと展望

本稿は、分類語彙表番号を付与したBCCWJと分類語彙表において多義語とされる語のリストを用い、相の類の多義語(多義的な用例分布の確認できた語)について、語義分布の実態調査を行った。また、多義の『日本国語大辞典』の出現順情報によって語義の新旧を確認したほか、先行研究が示す基本義や代表義の情報を確認し、出現頻度との異同を調査した。調査の結果、古い語義と代表義、高頻度語義が一致しない例が2/3を占め、いわゆるプロトタイプの語義を基本義とするならば、単純に基本義の判定はできないということが明らかとなった。

さらに、これらの情報と意味別の用例から、語義間の派生関係を分析し、派生関係情報の付与も行った。今後、用例を用いた語義の類似性判断実験を行い、派生関係のネットワークモデル化を図る。また、派生関係情報の付与が困難だった用例については、多義語に対する分類語彙表番号付与においては、判定揺れの可能性が高い。複数被験者による多義関係判断を取得することで、意味判定と意味分布の傾向も確かめる予定である。

謝辞

本研究は、2018年度国立国語研究所所長裁量経費・国立国語研究所コーパス開発センター共同研究プロジェクトおよび科研費

JP17H00917, JP18H05521, JP18K18519によるものです。

参考文献

Lakoff, George. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* The University of Chicago Press.

Langacker, Ronald W. (1987) *Foundations of cognitive grammar, Vol. 1. Theoretical prerequisites.* Stanford University Press.

Cruse, Alan. (2004). *Meaning in Language* (2nd Edition), Oxford University Press.

山崎誠・柏野和佳子(2017)『『分類語彙表』の多義語に対する代表義情報のアノテーション』『言語処理学会第23回年次大会発表論文集』, pp.302-305.

榎山洋介(2001)「多義語の複数の意味を統括するモデルと比喻」『認知言語学論考』1, pp.29-58, ひつじ書房.

榎山洋介(2002)『認知意味論のしくみ』研究社出版.

榎山洋介(2003)「多義性」松本曜編『認知意味論』pp.135-186, 大修館書店.

加藤祥・浅原正幸・山崎誠(2018)『『現代日本語書き言葉均衡コーパス』の新聞・書籍・雑誌データに対する分類語彙表番号付与』、『日本語学会2018年度秋季大会予稿集』, pp.161-166.

Sachi Kato, Masayuki Asahara, and Makoto Yamazaki. *Annotation of 'Word List by Semantic Principles' Labels for the Balanced Corpus of Contemporary Written Japanese, PACLIC-32.*

国立国語研究所(1964)『分類語彙表』, 秀英出版.
国立国語研究所(1984)『日本語教育のための基本語彙調査』, 秀英出版.

国立国語研究所(2004)『分類語彙表』増補改訂版, 大日本図書.