

令和元年6月12日現在

機関番号：13904

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K02727

研究課題名(和文) 複合動詞と類義表現を結ぶ意味特徴

研究課題名(英文) Corpus based analysis of semantic features of compound verbs and their synonymous expressions

研究代表者

神崎 享子 (Kanzaki, Kyoko)

豊橋技術科学大学・情報メディア基盤センター・特任准教授

研究者番号：00450693

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：本研究では、コーパスから句も含めて複合動詞の類義表現を半自動的に抽出、複合動詞の類義表現を観察し分析を行ったものである。類義表現抽出の対象にするのは、国立国語研究所の『複合動詞レキシコン』に収録されている約2700語の複合動詞である。そのうち『分類語彙表』と対応付けをしたのが、高頻度に使用される400語である。

本研究では400語の類義表現として自動抽出語精査した結果合計2690語の類義表現語句を得、『分類語彙表』の意味項目と約60%が対応付けられた。分類語彙表の意味項目の「2.用の類」の「作用」と「心」に最も多くの対応が見られた。複合動詞は感性的な表現とも深い関係があることがわかった。

研究成果の学術的意義や社会的意義

複合動詞は、日常我々の身近に使われる動詞であり、日本語では豊かに発達している。複合動詞は、二つの構成動詞が組み合わせられて一語になったもので、複合動詞内部の構成動詞の組み合わせを統語的、意味的に分析する研究が多かった。一方、複合動詞は二つの単純動詞の意味の足し算ではないことは知られているが、実際、複合動詞が文脈中で表す意味を客観的にとらえる研究はこれまでなかった。

本研究では大規模データに基づく複合動詞の調査・研究を行った。コーパスから複合動詞と類似度の高い語と句(表現とよぶ)を抽出、表現の妥当性を精査した後、類義表現リストを作成し、類義表現を分類語彙表の意味分布と照らし特徴をとらえた。

研究成果の概要(英文)：Japanese is known to be rich in compound verbs consisting of two verbs joined together. Recently the Japanese compound verb lexicon was constructed by National Institute for Japanese Language and Linguistics (NINJAL). However, it has no information on relation with synonymous words and phrases. We automatically extracted synonymous expressions of compound verbs from corpus which was “5 hundred million Japanese texts gathered from web.” Then we manually compile the list of synonymous expressions of about 2700 Japanese compound verbs by assessing the result. Then we chose 400 compound verbs frequently used in corpus, and made synonymous words and phrases correspond to Japanese thesaurus “Bunruigoihyo” published by NINJAL. As a result, the number of synonymous expressions that we obtained is 2690. Among them, 60% of the expressions belonged to semantic categories in “Bunruigoihyo.” Most expressions were belonging to two categories, “action and effect” and “mind.”

研究分野：日本語学(語彙意味論)

キーワード：複合動詞 類義表現 コーパス 語と句

様式 C-19、F-19-1、Z-19、CK-19 (共通)

1. 研究開始当初の背景

我々に身近な複合動詞は、日本語では豊かに発達しており語数も多い。「動詞(連用形)+動詞」で一語を構成する複合動詞については、これまで様々な観点から研究されてきた。複合動詞の構成動詞間の関係性を捉える記述的研究や語構造の解明を行う研究(斎藤 1992, 斎藤・石井 1997, 姫野 1999)、従来の複合動詞形成メカニズムの理論的研究に加え、データの観察による新たな知見を取り入れた研究も行われ(影山 1993, 2013 由本 2005, 2013 松本 1998, 2009)、さらにコーパスから構成動詞間の結合度や関係を求める定量的調査、研究(石川 2010, 山口 2013)なども進められてきた。また言語資源としての複合動詞レキシコンの整備(国立国語研究所)も行われ、多角的に複合動詞研究が展開しつつある。

複合動詞を分解的に観察し、構成動詞の統語的意味的情報によって複合動詞の意味を考えることが重要であることはこれまでの研究の成果で明らかであるが、それだけでなく、文脈の中で複合動詞の意味をとらえるとどのようなようになるのか関心をもった。

また、近年、大規模なコーパスを利用することが可能になったため、言語学でも今までより規模の大きいデータに基づいた語彙の調査・研究を行うという方向性も出てきた。そこで規模の大きいテキストデータから、複合動詞の意味をとらえるという考えに至った。

2. 研究の目的

複合動詞は動詞を組み合わせて表現するため、複合動詞の類義表現は語だけでなく句にまたがることも多い例えば

「握り締める」:「強く握る」「固く握る」「ぎゅっと握る」

「飲み歩く」:「飲みに出かける」「飲みに行く」「繰り出す」

「取り付ける」:語の意味としての類似性が高い類義語「設置する」「装着する」

文脈が重なれば類義として得られる表現「固定する」「はめ込む」「吊り下げる」

我々は複合動詞を使って曖昧な意味をより限定したり、逆に句で言い換えてわかりやすい表現にしたりすることができる。複合動詞の意味を考える場合、関連のある句の表現にもアクセスする必要があるが、これまで句表現との関連を考える研究は少なかった。

このように本研究では、コーパスから句も含めて複合動詞の類義表現をとらえ、複合動詞の意味について分析を行う。さらに、単純動詞一語では表現できないどのような意味を複合動詞がカバーしているかなども考察する必要がある。複合動詞の類義表現の中には、上記の例のように句への置き換えなどで表すものがあり、そのような複合動詞は対応する一語の動詞がないと考えられる。動詞意味体系の中で、複合動詞は単純動詞では表せないどのような意味を表現しているのかについてとらえる。

複合動詞には統語的複合動詞と語彙的複合動詞の別があるが(影山 1993)、本研究では国立国語研究所で公開している「複合動詞レキシコン」で扱われている「動詞+動詞」型語彙的複合動詞を対象にする。

3. 研究の方法

対象にする複合動詞は、国立国語研究所の『複合動詞レキシコン』に収録されている約 2700 語である。最初に、複合動詞の類義表現をデータから自動的に抽出し、さらに、半自動的に複合動詞ごとに類義表現を意味別に分類したリストを作成した。

類義表現語句を意味別に分類したリストは収録されている約 2700 語の複合動詞に対して作成したが、分析の際に『分類語彙表』と対応付けたのは、高頻度で使用される 400 語である。

手順について図示すると次のようになる。

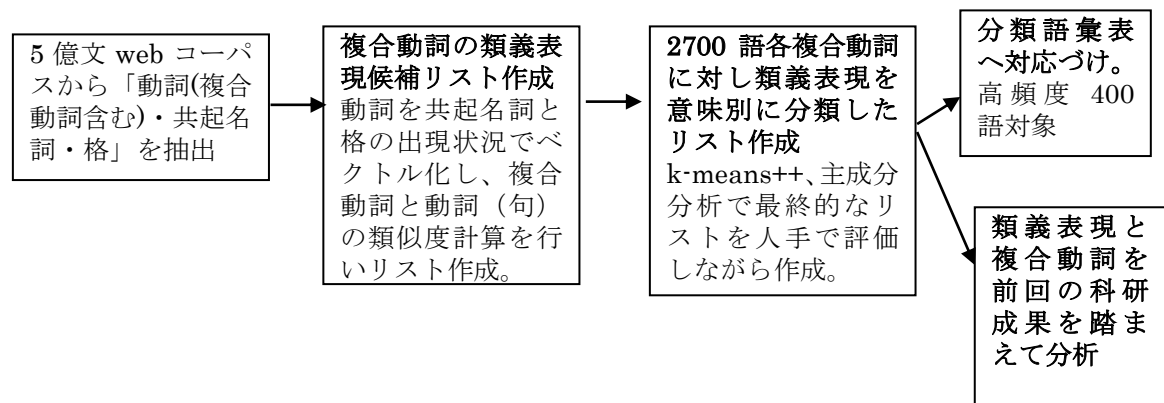


図1 類義表現リスト抽出手順

類義表現リスト作成については、まず5億文のコーパス(河原)から、辞書に登録されていない複合動詞をテキストから拾うための処理などの前処理をいくつか行ったあと、「すべての動詞(句)(複合動詞含む)・共起名詞・格」の組み合わせを取りだし、共起名詞と格の組み合わせから、複合動詞と動詞句の類似度を測った。その際、深層学習の word2vec でベクトル化し、ユザイン類似度という尺度で類似度を計算した。そして、この段階で得られた多義が含まれる

類義表現について、意味ごとに分類するために、クラスタリング手法 (k-means++) と主成分分析の手法を用いて、各複合動詞の類義表現語句の散布図を出力した。最終段階で、出力された散布図から人手でデータを精査し、最終的な類義表現リストを作成した。自動抽出した類義表現を人手で精査する際、評価基準となったのは『複合動詞レキシコン』に収録された意味項目である。『複合動詞レキシコン』には、各複合動詞に意味が記載されている。言い換え表現が類義表現として妥当かどうか、また、多義が合致しているか、あるいは計算機によって得た結果から新しい意味や表現が獲得されたかなどをチェックし、人手で取捨選択を行った。

4. 研究成果

本研究では高頻度に出現した 400 語の複合動詞の類義表現の異なりの数は、合計 2690 語、延べは 4511 語となった。複合動詞は句と関係が深く、たとえば「立ち寄る」は「ちょっと寄る」「ちらっと寄る」など副詞との関係や、「突き破る」は「破って飛び出す」のようにテ形連続の句の形などの例が多く見られた。これらの類義表現を『分類語彙表』の「2. 用の類」と対応付けた。句は『分類語彙表』に登録されていないため対応付けられなかったが、類義表現の中で、分類語彙表の意味体系と 1 対 1 で対応した語は 962 語/2690 語(35%)、類義語が複数の意味体系と一致しているのが、588 語/2690 語(21%)であった。現段階では句は対応付けられなかった。対応付けられなかった句は 1140 語/2690 語(42%)であった。

2.1(抽象的關係) 2.3(精神および行為) 2.5(自然現象)の大項目別では 2.1, 2.3 の大体の類義表現が分布している。さらに内訳をみると図 1 のようになる。

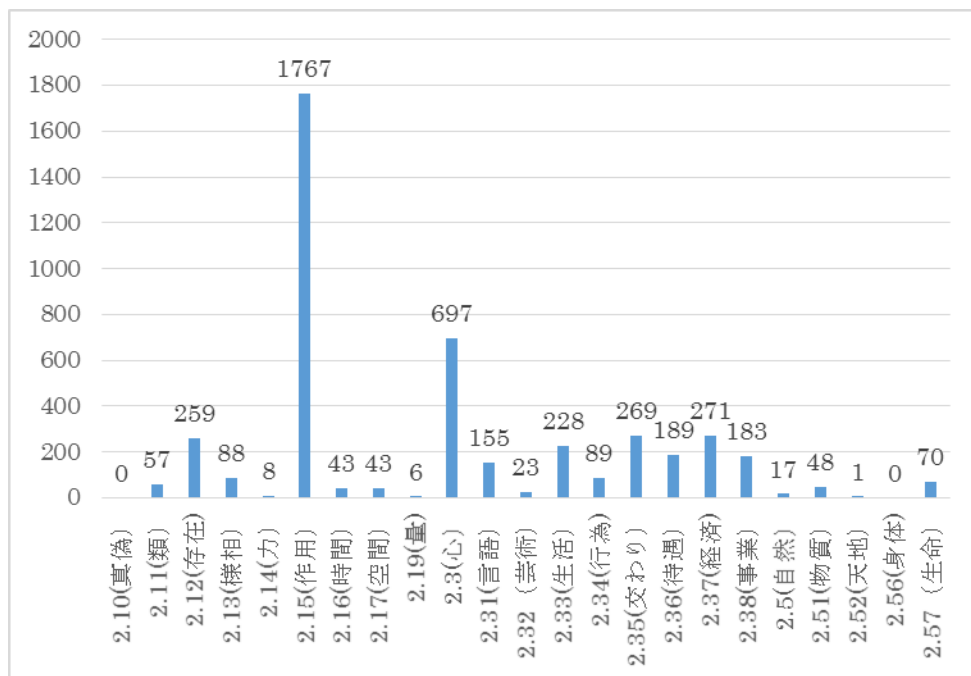


図 2 高頻度複合動詞の類義表現の意味項目別分布

分布をみると、2.15 の「作用」の意味項目に最も言い換え表現分布が多く、2.3「心」の意味項目がそれに続く。

2.15 の例としては、

「付け足す」

獲得した類義表現：

付け加える、補足する、追加する、足す、付加する [2.1580(関係/作用/増減・補充)]

さらに足す [登録なし]

などがみられる。2.3 の例としては、

「取り組む」

獲得した類義表現

尽力する、精励する、精進する、頑張る、勤しむなど [2.3040(活動/心/信念・努力・忍耐)]

などがみられる。

また、今回コーパスから作成した複合動詞の類義表現リストと、前回の科研の成果である統計的意味的情報を付与した構成動詞のデータとを合わせると、複合動詞と構成動詞に特徴がみられた。前項動詞 V1、後項動詞 V2 とする。① V1 の意味が反映され、V2 は副詞的に V1 の意味を一部重ねる形で意味合いをもたせる場合(「過ぎ去る」:(言い換え表現)とうに過ぎる、すでに経過する…)、② V1 の意味が主で、V2 は V1 の意味の一部と重なるが言い換え表現

中にV1の程度修飾表現が見られない場合(「願い出る」: (言い換え表現) 請願する、請う)、③分割できず一まとまりとして意味をもつ場合(「見下す」: 蔑む、侮蔑する)が観察された。これらは、主に意味的ヘッドがV1で、統語的継承性もV1の場合が多い。③以外は、構成動詞の意味が一部共通し構成動詞にも傾向が見られるため、V2から、結びつくV1を予測することができる。一方、言い換え表現の中に前項動詞連用形に助詞「て」がつく動詞連続の表現が見られる複合動詞があるが、その中にはいくつかのタイプがみられる。④両方の構成動詞が単純動詞そのままの意味ではなく、ニュアンスを帯びた意味合いになるもの(「持ち去る」: 奪って逃げる、盗んで逃げる、勝手に持ち出す)⑤V2の意味が中心で、V1の副詞的(連用的)な修飾でニュアンスが付与されるもの(「滑り落ちる」: 滑って落ちる、崩れて落ちる、力なく落ちる)⑥比較的V1の意味が中心でV2の意味が情報を付与すると同時に全体としてニュアンスが加わるもの(「飛び出す」: 飛んで出る、向かって突進する)⑦V2がV1の否定の意味を表すもの(「聞き落とす」)などがある、④~⑦はV1、V2の単純動詞の語彙の意味からは組み合わせが想定しにくい。傾向としては主に、意味的ヘッドも統語的継承もV2の場合はこのタイプが多い。

本研究成果により、従来なかった複合動詞の類義表現語句のリストを作成することができ、また、データを基盤にして、複合動詞を表現レベルで観察や考察を行うことができた。しかし、句と既存の意味体系を対応付けることは、句が登録されていないため、できなかった。国内外で語と句の接点をデータから分析したものは少ない。今後も、複合動詞の句も含めた表現を対象に、文脈との関係についてさらに発展させて考えていきたい。本研究で作成されたデータは、整備後、公開する予定である。

参考文献

- ①影山太郎『文法と語形成』ひつじ書房 1993,
- ②影山太郎編『複合動詞研究の最先端』ひつじ書房 2013,
- ③斎藤倫明『現代日本語の語構成論的研究: 語における形と意味』ひつじ書房 1992,
- ④斎藤倫明・石井正彦編『語構成』ひつじ書房 1997,
- ⑤石川慎一郎「現代日本語書き言葉均衡コーパス(BCCWJ)における複合動詞「~出す」の量的分析」『統計数理研究所レポート 238』統計数理研究所 2010,
- ⑥姫野昌子『複合動詞の構造と意味用法』ひつじ書房 1999,
- ⑦松本曜「日本語の語彙的複合動詞における動詞の組み合わせ」『言語研究 114』日本言語学会 1998,
- ⑧松本曜「複合動詞「~込む」「~去る」「~出す」と語彙的複合動詞のタイプ」,
- ⑨由本陽子・岸本秀樹編『語彙の意味と文法』くろしお出版 2009,
- ⑩山口昌也「複合動詞「~込む」と前項動詞の格関係」影山太郎編『複合動詞研究の最先端』ひつじ書房 2013,
- ⑪由本陽子『複合動詞・派生動詞の意味と統語』ひつじ書房 2005,
- ⑫由本陽子「語彙的複合動詞の生産性と2つの動詞の意味関係」影山太郎編『複合動詞研究の最先端』2013,
- ⑬国立国語研究所『複合動詞レキシコン』<http://vvlexicon.ninjal.ac.jp/>,
- ⑭国立国語研究所編『分類語彙表』増補改訂版 大日本図書 2004

5. 主な発表論文等

[雑誌論文] (計2件)

- ① Kyoko Kanzaki, Hitoshi Isahara, *Corpus based Analysis of Semantic Relations between Two Verbal Constituents in Lexical Compound Verbs*, proceedings of Asia Pacific Corpus Linguistics Conference (APCLC2018), 査読有り, 2018, 198-204

- ② Kyoko Kanzaki, Kenta Saito, Hitoshi Isahara, *Building a List of Synonymous Words and Phrases of Japanese Compound Verbs*, 11th edition of the Language Resources and Evaluation Conference (LREC2018), 査読有り, 2018, 2370-2373

[学会発表] (計2件)

- ① 神崎享子、井佐原均、コーパスにみられる語彙的複合動詞の語彙的結びつきと語用論的結びつき、The Japanese Society for Language Sciences 21st Annual International Conference (JSLs2019)、2019、査読有り、言語科学会(発表予定)

- ② 神崎享子、斎藤研太、井佐原均、語彙的複合動詞の類義表現抽出と多義別分類、言語処理学会第23回年次大会(NLP2017)、言語処理学会、2017、P16-2