

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 18 日現在

機関番号：13904

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520521

研究課題名(和文)複合動詞形成にみられる格の継承と意味のミスマッチ

研究課題名(英文)Generation of compound verbs observed from inheritance of case and semantic similarity between compound verbs and constituent verbs

研究代表者

神崎 享子(Kanzaki, kyoko)

豊橋技術科学大学・多言語翻訳研究本部・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) 2,500,000円

研究成果の概要(和文)：これまで複合動詞の言語分析には、定量的な調査を踏まえた研究が少なかったが、近年利用できるコーパスも大きくなり、統計的なデータと統語的な情報を用いた調査をすることができるようになった。そこで定量的に複合動詞と構成動詞の統語と意味のミスマッチを分析し、複合動詞形成のための構成動詞間の意味関係について研究を行った。具体的には、本研究では京都大学が公開している5億文のコーパスを使って、約2700語の複合動詞を対象にコサイン類似度で複合動詞が前項、後項動詞のどちらと類似度が高いかを計算して意味的類似性をとらえ、また複合動詞に対して格が複合動詞へどのように継承されているかを調査・分析した。

研究成果の概要(英文)：A quantitative analysis of Japanese compound verbs has ever been insufficient, however, since some huge corpora are released nowadays, a linguistics research can be conducted with such huge corpora and statistical data. For detecting objectively how to generate a compound verb from its two constituent verbs, I investigated an inheritance of syntactic cases and a semantic similarity between a compound verb and its constituent verbs. Similarity relations between 2700 compound verbs and their constituent verbs are calculated by cosine similarity measure based on co-occurring nouns in 5 billion web corpus. Considering six types of syntactic and semantic information, I found that two types of compound verbs exist. One is a combination of constituent verbs in lexical meaning level and another is a combination in context level. While a combination generated from context needs our general knowledge, a combination generated from lexical meanings can be analyzed systematically.

研究分野：日本語学

キーワード：複合動詞 意味関係 定量的調査

### 1. 研究開始当初の背景

「動詞 + 動詞」型の複合動詞は、東アジア言語に特徴的であり、特に日本語ではとても発達している。複合動詞のバリエーションが豊富で使用頻度も高いので、日本語学のみならず日本語教育や言語処理などの他分野でも、複合動詞研究は重要な課題の一つとなっている。理論的研究や記述的研究が行われてはいるが、扱うデータの数は限られており大規模な調査に基づいた研究には至っていなかった。コンピューターの普及や計算機上で扱えるコーパスのようなテキストデータが集積されてきている状況の中で、本研究は定量的調査によって統語的・意味的観点からシステムティックに複合動詞のしくみをとらえることを狙いとしている。

### 2. 研究の目的

本研究では、語彙的複合動詞形成の統語的・意味的情報をとらえるためのデータを作成し、その作成データを基盤に複合動詞形成の規則の解明を目指している。

和語の語彙的複合動詞は主に2語の動詞で構成されているが、その際、各構成動詞の本来もつ統語と意味が変化したり抑制されたりして一語の複合動詞を形成する。たとえば、「攻め上る」という複合動詞は、統語的にみると後項動詞「上る」の格関係が複合動詞へ受け継がれ、逆に前項動詞「攻める」のヲ格は抑制され、右側主要部の形になっている。しかし、コーパスで「攻め上る」の格に立つ名詞を調べると、「軍勢が都へ攻め上る」「デモ隊が国家の中核へ攻め上る」など、「攻める」の文脈に出現する名詞が多く、統語的な格と意味の継承にミスマッチが起きている。このような観察を通して、本研究では、  
(1) 統語と意味がどのように合成されているかについての分類、類型化  
(2) 統語と意味のミスマッチと複合動詞の意味との関係についての分析  
(3) 上記1, 2に基づいて、複合動詞化の意味的規則の解明を行う。

### 3. 研究の方法

対象となる複合動詞データは複合動詞レキシコン(国立国語研究所が公開)で対象となっている約2700語の語彙的複合動詞である。河原らが抽出した「webコーパス上の約5億文日本語テキスト」によって2700語の複合動詞の格関係と共起名詞を抽出した。具体的な方法は下記である。

(1) 京都大学が公開している形態素解析ツールJUMANと構文解析ツールKNPを利用して、5億文コーパスを構文解析した後、下記的情報を抽出した。

複合動詞の格および格に立つ名詞

各構成動詞(前項動詞と後項動詞)の格および格に立つ名詞

(2) 複合動詞と各構成動詞の類似度を、共起名詞の出現状況に基づいて求める。類似度計算するための類似尺度はコサイン類似度を用いた。類似度を計算した後、0~1の間に正規化した。これによって、共起名詞の出現状況からみると複合動詞が前項動詞と類似しているのか、後項動詞と類似しているのかを観察する。

(3) (2)では、どちらか一方の構成動詞が顕著に複合動詞と類似度が高ければ、その構成動詞は複合動詞と類似性があることを示し、もう一方が複合動詞と類似度が低ければ、その構成動詞は類似性が低いということになる。そこで、2つの構成動詞間で複合動詞との類似度の差をとり、顕著に片方の構成動詞が複合動詞と類似している場合について分析を行った。その際、閾値0.2を設定し0.2以上の差のあるものと0.2以下のものに分けた。

(4) 約2700語のリストには(1)~(3)で求められた意味的類似性の情報(前項・後項どちらの構成動詞が複合動詞と類似度が高いか)と統語的情報を付与した。統語的情報については、データからどちらの構成動詞がどのように複合動詞に統語情報を継承しているかを観察した。その結果、統語情報の継承の仕方には3種類のパターンがみられた。ここで前項動詞はV1、後項動詞はV2で表している。たとえば、「追い掛ける」であれば、統語的情報は前項動詞「追う」から継承し、「掛ける」の統語的情報は、「掛ける」が持つ二格が表層上には現れないことから、抑制されている。この場合は、統語的継承の種類としては「後項動詞の格の抑制(V1(-V2)と記す)」とした。また、たとえば、「願い出る」の場合、意味的類似性は前項動詞で、また統語的情報としては前項動詞のヲ格が付加される。この場合は、「前項動詞の格付加(V1(-V2)と記す)」とした。付加と抑制の違いは、複合化することにより、構成動詞がもっていた格が表層に現れない場合は「抑制」、両構成動詞の格が抑制されずに現れ、かつ、片方にしかない格も表層に現れた場合、「付加」とする。

また、片方の構成動詞の格がもう一方の格に交替した場合は「交替」とする。たとえば、「乗り付ける」では、前項「乗る」の二格に立つ名詞が、後項「付ける」のヲ格に立つ(車に乗る。ノ車を玄関に乗り付ける。)。このような場合を「交替」と分類した。

まとめると、

[統語的継承性]

前項動詞と後項動詞のどちらの統語情報を継承しているか。

どのように引き継いでいるか。タイプは3種類、「抑制」、「付加」、「交替」を観察した。

[意味的類似性]

前項と後項どちらの動詞が複合動詞と類似度が高いか。

このように、約 2700 語の複合動詞リストに意味的類似性と統語的継承性の情報を付与した。

#### 4. 研究成果

##### (1) 複合動詞のタイプ別統計

複合動詞を、構成動詞の統語的継承の観点から、次のような3つのタイプに分類した。

V1(-V2)：複合動詞は V1 の統語情報を継承

V2(-V1)：複合動詞は V2 の統語的情報を継承している。

V1, V2：複合動詞の文型が両構成動詞の文型と共通。

と については、3 節の(4)で述べたように、それぞれに「付加」「抑制」「交替」の3種類を区別した。

については、両構成動詞の文型が複合動詞の文型と一致する場合はほとんどであるが、両構成動詞からそれぞれ異なった格を複合動詞に継承している場合があった。

下記に、統語的継承のタイプと構成動詞間の類似度の差からみた複合動詞の内訳を示す。

	0.2 以上	0.2 未満	合計
V1(-V2)	173	289	462
V2(-1)	82	303	385
V1, V2	634	1259	1893
合計	889	1851	2740

表 1 統語的継承のタイプと構成動詞間の類似度の差からみた複合動詞の内訳

V1 と V2 の構成動詞間で類似度の差に開きがあるのは、複合動詞全体の 1/3 であり、残り 2/3 は差が比較的少ない 0.2 未満という結果になった。この結果から全体の 1/3 については、複合動詞に対する V1 と V2 の意味的優位性を考慮して複合動詞を分析できると考える。そこで、上記の表で V1 と V2 の差が 0.2 以上の複合動詞の中で、両構成動詞のどちらが複合動詞と類似度が高いか、内訳をカウントした。

V1(-V2)	170 語
V2(-V1)	53 語
V1, V2	481 語

表 2 V1 の類似度が高い複合動詞の数

V1(-V2)	3 語
V2(-V1)	29 語
V1, V2	153 語

表 3 V2 の類似度が高い複合動詞の数

理論言語学では統語的に右側主要部の規則が一般的であるが、後項動詞より前項動詞が統語的継承をする語が少なくなく、かつ、複合動詞の意味的類似度も後項動詞より前項動詞と類似度が高いということは、影山(2013)で述べられたように、日本語複合動詞の左側の語の役割を看過できないと考えられる。

また V1 と V2 の結合回数を調べたところ、V1 の類似度が高い場合と V2 の類似度が高い場合で、両者に共通してみられる傾向としては、結合回数の少ない動詞が全体の多くを占めているが、傾向として、複合動詞と類似度の高い構成動詞の方が、構成動詞の異なり数が多く、その代わり結合回数が少ない動詞も多い。類似度の低い構成動詞の方が、異なりが少ないが結合回数が多い傾向が見られる。複合動詞と類似度が低い結合回数が多い構成動詞の場合は、自立語の意味を変化させて接辞的な役割や動詞を補助する役割をもつ場合が含まれている可能性がある。

##### (2) 統語と意味のミスマッチと複合動詞の意味との関係

複合動詞との意味的類似性 V1、統語的継承性 V1(-V2)の場合

複合動詞は前項動詞の意味も統語も V1 から影響を受けているということになる。データの観察からは、複合動詞の中心的な意味は前項動詞が担っていると考えられる。

〔付加〕

複合動詞化しても自立語の場合と意味にずれが生じない場合、V1 が目的関係、あるいは原因関係の2つのタイプが観察された。たとえば「売り歩く」「待ちくたびれる(つかれる)」などである。「歩く」や「くたびれる(つかれる)」は、V2 の目的や原因という関係を持たば、柔軟に他の動詞を前項動詞にして結合でき、生産性がある。

V2 が自立語の意味用法とずれが生じている場合、V1 と V2 で共通する格の意味が V1 の意味を制限する。たとえば、「届け出る」の場合、ガ格と二格が共通の格で、ガ格が重なっていることから、自ら進んでという主体からの意志的なアクションを示し、二格が重なっていることから、行為の着点である対相手が想定される。また、「お弁当を届け出る」とは言わないことから、「出る」と結合することで、「届け出る」が「届ける」より意味が限定されていることがわかる。「届け出る」は、自分より社会的立場が上の人間あるいは社会的組織のところに現れて届けるという意味に限定されるようにみられる。「届け出る」と同じパターンのもので、「願い出る」「申し出る」「迎え出る」などがある。他に「離れ去る」「逃げ去る」「書きなぐる」「言い捨てる」なども〔付加〕のタイプになる。

〔交替〕

V2 の格は表面上にはあらわれていないが、V1 の「コト」を受けていて補文関係となる。たとえば、「読みふける」「見慣れる」「睨み返す」などがこの類である。

〔抑制〕

V1 と V2 で格が共通の意味を V1 に付加しているという点では〔付加〕と同様である。たとえば、「こぼれ出す」「流れ出す」はヲ格が抑制されていて、また、起点や着点を表すカラ格と二格が共通。V1 の方向性の意味が強められ

ている。  
V2の格を抑制しているという点に〔付加〕との違いがみられる。格を抑制していることから、V2の意味は自立語の意味から意味を変化させてV1を修飾していると考えられる。V2とV1の結合は、統語的複合動詞のように比較的自由的な結合を許すのではなく、ある程度特定される。このタイプでは、V2はV1の語彙的意味の一部を修飾し、方向性を強めたり、程度を強めたり、副詞的にV1の様態を修飾するなどがみられる。たとえば、「聞き入れる」「決め付ける」「書きなぐる」「読み返す」「読み流す」などはこの類である。

複合動詞との意味的類似性 V2、統語的継承性 V1(-V2)の場合

このタイプでV1とV2の差が大きいデータ、つまり、どちらか一方の複合動詞と類似度が高いという複合動詞は「連れ帰る」「移り住む」の2例だけであった。「V1でV2」の時系列や付帯状況なども表す関係と想定される。

複合動詞との意味的類似性 V1、統語的継承性 V2(-V1)の場合

意味はV1と類似しているのが、統語的な継承はV2から行われるというタイプである。

〔付加〕

V2はV1の語彙的な要素を修飾するのではなく、文脈的結びつきになる。つまりV1の行為とV2の行為を組み合わせている。V2からみればV1がV2の仕方、方法を表す。たとえば、「酌み交わす」では相手格のト格が継承され、ヲ格は共通である。たとえば「酒を酌んで交わす」という例になる。他には「飲み比べる」「見分ける」「抱え出す」「履き替える」「割り入れる」「鳴り渡る」などがある。

〔交替〕

V1の格をV2の格に変えることで、V2の意味が付け加わる。たとえば、「噛み付く」では、「噛む」のヲ格の対象を付く「二格」に変えることで噛むという行為に、主体の相手への方向が強められ、主体の動きが加えられる。他に「乗り入れる」「酔いつぶす」「殴りかかる」「追いつがる」などがある。

〔抑制〕

V1がおもな意味を担っているが一部格情報が抑制されている。またV2によってV1の行為に動きを加えている。「這い寄る」「書き飛ばす」などがこの類である。

複合動詞との意味的類似性 V2、統語的継承性 V2(-V1)の場合

〔付加〕

V1でV2時系列、因果の場合もある。最終的にはV2が終点の行為である。

泣き腫らす、取り替える。踏み入る

〔交替〕

因果関係になるような類似した表現を重ね合わせている。どのような状態でV2するか、V2の仕方をV1表す。たとえば、「漏れ聞く」「突き刺さる」「覆いかぶさる」などがあげられる。

〔抑制〕

V1は接辞的に意味を添えており、V2の語彙的意味を強めている。

引き締まる、引き籠もる 突き入る、突き進む、分け入る、打ち震える、掻き曇る

(5)まとめ

本研究では、複合動詞と構成動詞の類似性と統語的継承の情報を付与した複合動詞データを作成した。そして、特にどちらか一方の構成動詞が複合動詞と高い類似性を持っており、かつ統語的にもどちらか一方の格を複合動詞に継承している場合について、データに基づいて分析した。

複合動詞の合成の仕方には大きく二つのタイプがあり、V1とV2の語彙的意味の内部的な関係によって複合動詞が合成されるものと、語彙的意味それ自体の変化はなく、文脈的にある行為どうしが結びつき複合動詞が合成されているものがある。

文脈的に行為動詞が結びついているタイプは次のようなものである。類似性はV1、格の継承はV1(-V2)のタイプでは、〔付加〕の「売り歩く」「待ちくたびれる」のように、V2の「歩く」や「くたびれる/疲れる」と目的や原因結果の関係をもてば、V1の語は特化されずに広くいろいろな動詞が結びつくことができる。また格の継承はV1からであっても(V1(-V2))複合動詞との類似性がV2の場合、時系列あるいは付帯状況の関係で動詞が結合する。「連れ帰る」「移り住む」などがそのタイプである。格の継承がV2(-V1)の場合はV1とV2が文脈的結合になることが多い。また、複合動詞との類似がV1の場合、〔付加〕がそのタイプである。一方、複合動詞の類似がV2の場合、〔付加〕と〔交替〕が担っている。

上記以外が、語彙的意味の内部で変化や修飾関係をもって複合動詞を合成しており、そのようなデータが多い。

本研究ではタイプの詳細な分類により動詞と動詞が意味を変化させず文脈的結びつきをしているものと、語彙的意味のレベルでの意味関係で合成されているものを区別することができた。文脈的結びつきのものは、因果を表したり目的を表したり、V2の仕方などを表せば、柔軟に複合動詞を形成することができるが、しかし反面、因果関係などは知識や常識によって結びつくことになるので、語彙の意味記述レベルで静的にすべての可能性を記述しておくことができない。

その一方で語彙的意味を変化させている場合は、システムティックに複合動詞の意味的な合成の仕方について記述することができる。分類ごとに各複合動詞について語彙的意味の合成の仕方について分析を行った。

(参考文献)

影山太郎(2013)「語彙的複合動詞の新体系」  
影山太郎編『複合動詞の謎に迫る』pp.3-46  
ひつじ書房

5. 主な発表論文等  
(研究代表者、研究分担者及び連携研究者には下線)

研究者番号：

〔雑誌論文〕(計 件)

〔学会発表〕(計 2 件)

Kyoko Kanzaki, Compound Verb Lexicon,  
日中自然言語処理共同促進会議  
(CJNLP2014), 2014年10月12日~14日, 成  
都(中国)

Hitoshi Isahara and Kyoko Kanzaki,  
Parallel Data Sharing for  
Intensification of the International  
Competitiveness, MAPLEX2014, 2015年2  
月9日~11日, 天童(山形)

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
取得年月日：  
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 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 准教授  
研究者番号：00450693

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )