

現代日本語における汎用的漢語サ変動詞の抽出 とその内部構成の検討

李 楓 (神戸大学大学院国際文化学研究科)

Identification of Major Sino-Japanese-Verbs Ending with *-suru* and Analysis of Their Internal Structures

Feng Li (Graduate School of Intercultural Studies, Kobe University)

1. はじめに

世界の中で日本語教育が広く行われるようになるにつれ、日本語の語彙指導についても関心が高まっている。日本語の語彙には、和語、漢語、外来語など、様々な語種が存在するが、数のうえで特に多数を占めるのは漢語であり、その多くが漢語サ変動詞の形を取るとされている。

しかしながら、外国人日本語学習者、特に学習者数の多い中国人日本語学習者の視点で見した場合、漢語サ変動詞の種類や用法などについて、必要とされる情報がすべて明らかにされているとは言えない状況にある。たとえば、国語辞典や日本語語彙教材では、漢語名詞の後に「～する」と添え書きしたり、漢語名詞の用例の一部として「～する」の形を示したりするのが一般的で、漢語サ変動詞そのものを単独の語として認定し、その使い方を体系的に説明しているものはほとんど存在しない。また、無数に存在する漢語サ変動詞の中で、典型的に使用される漢語サ変動詞にどのようなものがあり、それらがどのような内部構成を持っているかといった点についてもはっきりしない。

もちろん、日本語学の中では、漢語サ変動詞の用法に関わる問題は古くより取り上げられてきた。ただし、先行研究の多くはもっぱら理論的な分類や構造の解明を目指しており、現代日本語における実際の使用状況をふまえた研究は必ずしも多くない。しかしながら、学習者のニーズから言うと、まずもって明らかにされるべきことは、日本語を構成する様々な変種において広く使用されている漢語サ変動詞にはどのようなものがあり、それらがどのような内部構成パターンを持つのかのかわかりやすく示されることであろう。

以上の点をふまえ、本研究においては、日本語初的大型コーパスである現代日本語書き言葉均衡コーパス (BCCWJ) を用い、各種の日本語変種の中で、汎用的、かつ、高頻度で使用されている漢語サ変動詞を特定したうえで、その内部構成を概観していくこととした。BCCWJ には、様々な言語データが収められているが、本研究では、一定数の読者を持ち、外国人学習者にとっても重要であると考えられる書籍、雑誌、新聞、ブログ、白書、知恵袋の6種に限り、頻度調査を行っていくこととする。

2. 先行研究

すでに述べたように、本研究では、現代日本語の各変種の頻度をふまえ、汎用的な漢語サ変動詞を特定すること、及びその内部構成の特性を明らかにすることを主たる目的とする。前者に関わる先行研究としては、日本語の計量的な語彙調査が古くより活発に行われている。例えば、国立国語研究所 (1962 ; 1973 ; 1983) では、雑誌、新聞、教科書などの語彙調査が行われており、漢語を含む日本語語彙の使用状況が一定の範囲で明らかにされている。また、橋本 (2009) や松下 (2011) は、BCCWJ を用いて、教育的語彙リストの作成を試みている。しかしながら、漢語サ変動詞を対象を絞り、また現代日本語の多様な変種に目配りを行った調査は必ずしも十分ではない。

後者については、特に2字漢語から派生した漢語サ変動詞について、理論的立場から、内部構成の分類方法がいくつか提唱されている。ここでは、代表的なものとして、日向 (1985)、野村 (1999)、小林 (2004) を概観する。

日向(1985)は、現代日本語における漢語サ変動詞の構造を、並立関係(添加、開閉など)、修飾関係(直送、軽視など)、客体関係(加熱、失望など)、実質関係(強化、酸化など)の4種類に分類している。同研究では、4分類のうち、特に修飾関係と客体関係に焦点を当てて考察している。

野村(1999)は、漢語サ変動詞の分類を精緻化するため、まず、漢語を構成する個々の文字タイプと、文字間の基本的な結合タイプをそれぞれ5種類に下位分類したうえで、これらを組み合わせ、全体で22種類からなる内部構造パターンを提唱している。文字タイプの下位分類は、叙述の対象となる物や事を表す「事物類(N)」(鉄、国、水、土、道など)、事物の動作・作用を表す「動態類(V)」(見、増、置、感など)、事物や精神の性質・状態を表す「様相類(A)」(新、軽、大、高など)、動作や状態の程度・内容を限定・修飾する「副用類(M)」(特、再、絶、予など)、語基について形式的な意味を添える「接辞(s)」(不、御、的、性など)の5種類である。次に、結合タイプの下位分類は、補足関係(+)、修飾関係(>)、並立関係(・)、対立関係(-)、反復関係(=)の5種類である。これらを組み合わせることで、漢語サ変動詞は、「N+V」(気絶、骨折など)、「V+N」(握手、開花など)、「A+N」(多言、貧血など)、「N>V」(音読、兄事など)、「V>V」(愛用、滑降など)、「A>V」(安眠、軽視など)、「M>V」(一掃、共感など)、「V・V」(引退、救助など)、「V-V」(開閉、屈伸など)、「V=V」(云々、転々など)、「N>N」(金策、原因など)、「V>N」(起因、残業など)、「A>N」(紅葉、粗食など)、「N・N」(影響、葛藤など)、「N-N」(左右、始末など)、「A=A」(清々)、「sV」(殺到、所期など)、「sA」(不精)、「Ns」(液化、酸化など)、「Vs」(欠如、消化など)、「As」(悪化、強化など)、「その他」(運休、軍縮など)の22種に整理される。

小林(2004)は、野村(1999)をはじめとする一連の研究を継承しつつ、漢語サ変動詞の範囲を漢語部分が2字以外のものに拡張し、幅広い漢語の構成を分析した。このうち、2字漢語(「動名詞」と称される)については、動詞的要素と名詞的要素で構成されるVN-Nタイプ(読書、投票など)、動詞的要素と動詞的要素で構成されるVN-VNタイプ(使用、殴殺など)、付加詞的要素と動詞的要素で構成されるADJ-VNタイプ(銃殺、病死など)、及び構成要素が抽出できないもの(挨拶、支配、勉強など)の4種類に分類している。なお、VN-Nタイプについては、内部の名詞的要素と関係づけられた項を取るかどうかという観点から、項を取れないタイプ(飲酒、処刑など)、項を取れるタイプ(投票、登山など)、項を取らなければならないタイプ(開封、除名など)の3種類に下位分類している。また、VN-VNタイプについては、語の意味の中心をなし、語全体の品詞を決める主要部の性質や構成要素間の意味的結合関係に基づき、両側主要部タイプ(使用、委託など)、右側主要部タイプ(殴殺、急行など)、左側主要部タイプ(採用、購読など)の3種類に下位分類している。

以上の3氏の分類を整理すると、おおよそ以下のような対応関係になると考えられる(表1)。ただし、この対応関係は大まかなもので、必ずしもこのように明示的に述べられているわけではないことに注意されたい。

表1 先行研究における漢語サ変動詞の内部構成

日向(1985)	並立関係	修飾関係	客体関係	実質関係	
野村(1999)	V・V、V-V、 V=V、N・N、 N-N、A=A	A+N、N>V、 V>V、A>V、 M>V、N>N、 V>N、A>N	V+N、 N+V	sV、sA、 Ns、Vs、 As	その他
小林(2004)	VN-VN	ADJ-VN	VN-N	構成要素が抽出できないもの	

以上で、日向(1985)、野村(1999)、小林(2004)の3つの枠組みにおける漢語サ変動詞の内部構成パターンの分類方法について概観してきた。これらを概観して気が付くことは、

3氏はそれぞれ類似した方向性を持ちながらも、細かい点では異なりが多いことである。たとえば、内部構成パタンの数は、日向(1985)では4種類、野村(1999)では(基本タイプに限ると)5種類、小林(2004)では主として3種類である。また、構成要素間の結合関係パターンに注目すると、日向(1985)が提唱した客体関係と実質関係は野村(1999)では採用されておらず、一方、野村(1999)は新たに補足関係、対立関係、反復関係を加えている。これに対し、小林(2004)はもっぱら統語的關係性に分類の焦点を当てており、結合関係パターンについては明確に論じられていない。

3氏の分類はそれぞれに理由のあるものであると考えられるが、教育的観点から漢語サ変動詞の内部構成のパターンを示そうとする場合、これらの先行モデルの統合と精選が必要になるだろう。

3. リサーチデザイン

3.1 本研究の目的

既に述べたように、本研究は、書籍、雑誌、新聞、ブログ、白書、知恵袋の6種の変種データを用い、汎用的漢語サ変動詞を計量的に特定したうえで、その内部構成を明らかにすることを目的とする。調査にあたり、以下のリサーチクエスチョンを設定した。

RQ1. 現代日本語の各変種において、高頻度漢語サ変動詞に内容的な違いが見られるか？

RQ2. 各変種の頻度情報を統合することで、どのようなものが汎用的漢語サ変動詞として抽出されるか？

RQ3. 高頻度・汎用的漢語サ変動詞の内部構成パターンとしては、どのようなものが多いか？

3.2 データ

BCCWJで使用する書籍、雑誌、新聞、ブログ、白書、知恵袋の6変種のうち、書籍については、「出版・書籍」と「図書館・書籍」を分析対象とする。前者は、2001年から2005年の間に国内で刊行された書籍で、国立国会図書館の蔵書目録を電子化した「J-BISC」を元に決定された母集団から、日本十進分類法(NDC)及び発行年度ごとに層別に抽出したものである。後者は、東京都立中央図書館の作成した「ISBN総合目録」に基づき、1986年から2005年までの20年間に発行された書籍を母集団としたランダムサンプルである。「収集目的が異なるが、2種類の書籍データはともに、NDCと発行年をもとにランダムサンプリングされているため、任意に抽出した部分も母集団のNDCの構成比をそのまま表していることが期待される」(大石、2012)ことから、ここでは、2種類の書籍データをまとめて使用する。

ただし、書籍はBCCWJの中で、圧倒的なデータ量を占めている。そこで、以下の分析においては、書籍を1変種として扱うのではなく、NDCによる10ジャンル(0.総記、1.哲学、2.歴史、3.社会科学、4.自然科学、5.技術・工学、6.産業、7.芸術、8.言語、9.文学)をそれぞれ独立した言語変種と見なして分析を行う。これにより、他の変種との分量バランスの偏りが改善される。すなわち、本研究で、対象とするのは、書籍の10種、その他の5種、あわせて15種の変種データである。

漢語サ変動詞の頻度情報の取得には、「中納言」を使用する。「中納言」では、短単位検索、長単位検索、文字列検索の3種類の検索が可能であるが、本研究の目的に応じて、短単位検索を利用することとした。データの採集は2013年7月に実施した。

3.3 調査対象の定義

小林(2004)のように、漢語サ変動詞を広義で捉えれば、1文字のもの、2文字のもの、3文字以上のものなど、幅広い語が対象に含まれる。しかし、本研究においては、2文字からなる漢語サ変動詞を対象を限定する。これは以下の3つの理由による。1点目として、種類や使用頻度の点において、2字漢語が漢語語彙の圧倒的多数を占めるためである。2点目

は、3 字以上の漢語に比べて、2 字漢語は「ひとまとまり性」が強く（湯本、1977）、計量的に調査しやすいためである。3 点目は、中日 2 言語の対照性が高く、中国人学習者にとって馴染みの深い語が多いためである。

しかし、対象を 2 字漢語と定めても、具体的に用例を検証していくと、多くの境界例が存在することに気が付く。そこで、統一した処理を行うために、本研究で対象とする漢語サ変動詞の決定ルールを以下のように定めた。

- (1) 「X する」（「する」の活用形を含める）の形が存在する。
- (2) 「X する」が全体として何らかの動作もしくは変化を表す。
- (3) 「X」は原則として 2 文字とする。ただし、「X」が 3 文字以上の場合、後項 2 文字だけで独立した意味を持ち、かつ、「する」と結合しうる場合は、「後項 2 文字+する」の形で対象に含める。

対象になるものと対象にならないものの例を以下の表 2 に示す。

表 2 本研究が対象とする漢語サ変動詞の例

	対象となるもの	対象とならないもの
「X」が 2 文字	感謝する、 <u>1</u> 泊する、 <u>2</u> 分する	艶々する
「X」が 3 文字	再確認する（→確認する）、一安心する（→安心する）	不自由する、土下座する
「X」が 4～文字	調査検討する（→検討する）、悪戦苦闘する（→苦闘する）	右往左往する、意気投合する

「1 泊する」や「2 分する」は漢数字の異表記と見なして対象に含める。一方、「艶々する」は形容詞的に状態を描写し、上記のルール (2) に抵触するため、対象外となる。また、「不自由する」や「右往左往する」などは、後項 2 文字部だけで「する」と結合することではなく、ルール (3) に抵触するため、同じく対象外となる。

3.4 手法

まず、RQ1 については、15 種の言語変種ごとに、「中納言」を用いて漢語サ変動詞の抽出を行う。その際、検索対象を「スル（語彙素読み）」に指定する。出力された結果をすべてダウンロードし、前述のルールと照合して、対象語の特定を行う。

なお、「X する」の「X」部分が 3 文字以上からなるものについては、前節のルール (3) により、「後項 2 文字+する」の形が成立すれば対象に含めるわけだが、当該形が日本語として成立するかどうかの判断は時に主観的となりうる。そこで、当該形がコーパス内で 2 例以上出現していることを当該形の成立の判断根拠とする。「不自由する」の例で言うと、「後項 2 文字+する」形である「自由する」が仮にコーパス内で 2 例以上存在していれば、対象に含めることになる（2 例以上とすることで、特殊な効果を狙った例外的用例などが排除できる）。その後、言語変種別に高頻度に使われている漢語サ変動詞をリストアップし、サンプルとして上位 5 語を概観する。

次に RQ2 については、まず、15 種の言語変種からそれぞれ高頻度上位 50 語、延べ 1250 語を抽出し、重複を除いた 279 語を決める。次に、一定の汎用性を確認するため、15 変種中、1 変種でしか上位 50 語に入っていない 2 語（「出走」、「共起」）を除き、277 語を対象として変種別の頻度調査を行う。その後、主成分分析により、15 変種中の頻度を合成する。最後に、第 1 主成分得点に基づいて 277 語を降順に並べ替え、主成分得点が正（+）となる 93 語を高頻度かつ汎用的な漢語サ変動詞として特定する。

最後に、RQ3 については、RQ2 で特定された 93 語の各々を次節で述べる内部構成パターン別に分類する。その後、タイプベースとトークンベースの 2 種類の観点で計量的に概観し、

最も典型的な構成パターンを明らかにする。タイプ（異なり語数）ベースでは、分析対象語の総語種数（93）を分母として、各パターンに含まれる語種数比率を調べる。また、トークン（延べ語数）ベースでは、93語の各々が持つ第1主成分得点の総計値を分母として、各パターンに含まれる語の第1主成分得点の合計値の比率を求める。

3.5 内部構成パターンの分類法

既に述べたように、漢語サ変動詞の内部構成モデルを提唱した日向（1985）、野村（1999）、小林（2004）の3氏の枠組みの間には微妙なずれがあり、また一部の分類は過剰に細かなもので、学習者にとってわかりやすいものとは言い難い。

そこで、本研究では、先行研究の分類の融合と精選を行う。

まず、構成要素間の結合関係パターンについては、日向（1985）の枠組みに新たに「補充関係」を加え、全体で5分類とする。「補充関係」というのは、「拡大」「縮小」などのように、後項構成要素が前項構成要素の意味を補充するものである。また、品詞的結合パターンについては、野村（1999）で提唱されている22分類のうち、「A=A」、「sA」及び「その他」を削除し、「V・V」、「V-V」、「V=V」の3種を「V-V」、「A+N」と「A>N」の2種を「A>N」にまとめ、さらに「M・s」を加え、合計18分類とする。「A=A」、「sA」を除外したのは、当該例における語幹部の名詞性が曖昧であると考えたためである。なお、品詞的結合関係を表す記号については、野村（1999）を参考にしながら、並立関係は「-」、修飾関係は「>」、補充関係は「<」、実質関係は「・」とする。以上により、本研究で使用する分類の枠組みは以下の表3に示すとおりとなる。例の大部分は野村（1999）から引用している。

表3 内部構成の枠組み

構成要素間 結合関係 パターン	品詞的結合 パターン	例	構成要素間 結合関係 パターン	品詞的結合 パターン	例
並立関係	V-V	選択、開閉	客体関係	N+V	骨折、気絶
	N-N	意味、意見		V+N	看病、入院
	A>N	紅葉、大病	補充関係	V<A	拡大、縮小
	A>V	短縮、軽減		V・s	進化、激化
修飾関係	M>V	予防、再建		N・s	風化、電化
	N>N	手術、病気	実質関係	A・s	酸化、強化
	V>V	誤診、傾聴		M・s	特化
	N>V	林立、列举		s・A	不精
	V>N	起因、残業		s・V	否認、不足

漢語サ変動詞の構成要素となる個々の文字の意味や品詞性については、判断の揺らぎを避けるため、『新潮日本語漢字辞典』（新潮社、2008）、『漢字源 改訂第五版』（学習研究社、2011）、『三省堂常用漢字辞典』（三省堂、2013）の3種類の漢字辞書の記述に従い、当該文字の持つ訓や意味を基準として分類する。1つの漢字が2つ以上の訓読みや意味を持つ場合、また、同じ意味を表す訓が複数ある場合などは、機械的に最も先頭に書かれているものを採用する。また、辞書間で記述が揃っていない場合は、3つの辞書のうち、2つに出現しているものを採用する。さらに、3辞書間で統一した意味が見つからない場合は、収録内容が最も多く、記述が最も詳細である『新潮日本語漢字辞典』に準ずる。

以上のような処理手順を取ることで、同じ語が先行研究と異なる分類を与えられる場合もある。例えば、「挨拶」を例にすると、小林（2004）はこれを「構成要素が抽出できないもの」としているが、上記の辞書の解説をふまえると、「挨拶」は「^{せま}挨拶+^{せま}する」と読み下ろすことができるので、本研究ではこれを並立関係の「V-V」と認定する。なお、このとき、「挨拶」という語の意味と「挨拶／する」（相手に近付く）の意味の間には若干のず

れがあるが、本研究では、構成要素の意味間の関係性を基準として分類する。

4. 結果と考察

4.1 RQ1 日本語の各変種における高頻度漢語サ変動詞

15 種の言語変種ごとに高頻度漢語サ変動詞を調査したところ、以下の表 4 のようになった(語幹部のみを示す)。

表 4 言語変種ごとの高頻度漢語サ変動詞(上位 5 語)

表1：言語変種としての同類及供語 / 変動詞（主語：語）					
書籍 ・総記		書籍 ・哲学	書籍 ・歴史	書籍 ・社会科学	書籍 ・自然科学
1	表示	存在	発見	存在	存在
2	利用	理解	存在	実施	利用
3	実行	説明	主張	利用	発生
4	使用	実現	出土	説明	結合
5	発生	意味	利用	規定	報告
書籍 ・技術・工学		書籍 ・産業	書籍 ・芸術	書籍 ・言語	書籍 ・文学
1	表示	利用	演奏	存在	説明
2	使用	提供	表現	説明	結婚
3	利用	紹介	紹介	意味	発見
4	発生	発生	存在	表現	心配
5	設定	使用	使用	注意	存在
雑誌		新聞	ブログ	白書	知恵袋
1	発売	発見	発売	実施	表示
2	開催	期待	開催	開催	削除
3	紹介	予想	紹介	設置	質問
4	用意	確認	確認	期待	使用
5	掲載	注目	期待	決定	購入

上記で明らかなことは、変種により、高頻度漢語サ変動詞の内容に大きな違いが存在することである。例えば、15 変種において、共通して上位 5 語に含まれた語は 1 語も存在しない。このことは、現代日本語における漢語サ変動詞を議論するにあたり、各種の言語変種の差に十分に留意する必要があることを示している。

4.2 RQ2 高頻度・汎用的漢語サ変動詞の特定

前述の 277 語について 15 種の変種ごとに頻度を調査し、主成分分析を実行したところ、一般に有効主成分とされる固有値 1.0 以上の主成分が 5 つ抽出された。

このうち、第 1 主成分に注目すると、15 変種すべてに対して負荷量が正となっており、第 1 主成分が複数変数の合成指標となっていることが確認された。第 1 主成分の寄与率は 38.33% となり、元のデータの分散の約 4 割が第 1 主成分に集約されたことになる。以下に示すのは、15 変種の第 1、第 2 主成分負荷量表(表 5)、それらを横軸・縦軸とする二次元象限上にデータを布置した散布図(図 1)、及び、第 1 主成分得点の降順で全体を並べ替え、指標値が正になった 93 語のリスト(表 6)である。

はじめに、表 5 と図 1 に注目する。15 変種の中では、書籍、特に社会科学(0.816)や産業(0.809)の分野に相対的に大きな負荷量がかかっている。しかし、ブログ(0.521)など、書籍以外の変種にも一定の負荷量がかかっており、全体として日本語の多様な言語変種をバランスよく代表した指標値が取り出せたと考えられる。

表5 主成分負荷量

変種	主成分 1	主成分 2
書籍・総記	0.571	0.313
書籍・哲学	0.704	-0.521
書籍・歴史	0.747	-0.277
書籍・社会科学	0.816	-0.140
書籍・自然科学	0.644	-0.180
書籍・技術・工学	0.742	0.250
書籍・産業	0.809	0.025
書籍・芸術	0.748	-0.101
書籍・言語	0.556	-0.506
書籍・文学	0.575	-0.323
雑誌	0.525	0.535
新聞	0.368	0.361
ブログ	0.521	0.589
白書	0.327	0.332
知恵袋	0.307	0.429

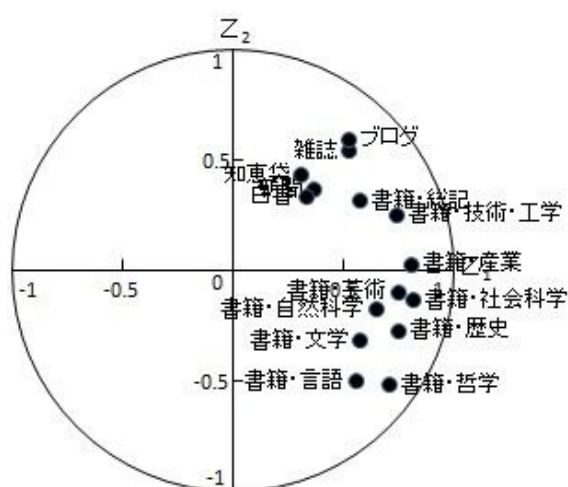


図1 負荷量散布図

表6 高頻度・汎用的漢語サ変動詞 (主成分得点>0)

順位	語	得点	順位	語	得点	順位	語	得点	順位	語	得点
1	存在	12.457	25	実現	3.353	49	発売	1.841	73	発揮	0.671
2	利用	11.823	26	表現	3.207	50	意識	1.806	74	一致	0.626
3	紹介	8.992	27	展開	3.064	51	移動	1.728	75	無視	0.621
4	説明	8.984	28	参加	3.058	52	指定	1.713	76	維持	0.534
5	使用	8.711	29	検討	3.023	53	強調	1.703	77	適用	0.506
6	確認	8.327	30	意味	2.827	54	発展	1.546	78	確立	0.506
7	理解	7.135	31	対応	2.823	55	増加	1.521	79	記載	0.499
8	表示	6.141	32	主張	2.797	56	開発	1.467	80	勉強	0.469
9	発見	5.946	33	要求	2.787	57	否定	1.402	81	分類	0.451
10	発生	5.726	34	完成	2.649	58	提出	1.368	82	安定	0.368
11	注目	5.148	35	判断	2.642	59	反映	1.332	83	観察	0.360
12	指摘	4.992	36	成功	2.609	60	規定	1.239	84	保存	0.303
13	期待	4.906	37	設置	2.606	61	代表	1.095	85	心配	0.281
14	実施	4.408	38	決定	2.431	62	掲載	1.024	86	重視	0.264
15	用意	4.339	39	作成	2.350	63	支配	1.013	87	考慮	0.253
16	発表	4.018	40	選択	2.208	64	結婚	0.960	88	販売	0.220
17	注意	3.989	41	開始	2.183	65	実行	0.884	89	想像	0.209
18	開催	3.834	42	採用	2.162	66	変更	0.858	90	活用	0.140
19	構成	3.658	43	登場	2.135	67	拡大	0.774	91	限定	0.133
20	評価	3.649	44	認識	2.046	68	解決	0.768	92	減少	0.129
21	設定	3.623	45	予想	1.984	69	解放	0.764	93	区別	0.111
22	変化	3.488	46	成立	1.974	70	比較	0.734			
23	提供	3.436	47	報告	1.901	71	集中	0.727			
24	形成	3.365	48	導入	1.873	72	購入	0.683			

次に、表6に注目する。一般に、雑誌や新聞など、情報伝達を主とする言語変種に限定して頻度調査を行えば、当該変種の特性に影響を受け、分野に依存した難語のみが多く選定されがちである。しかし、幅広い言語変種をデータに加えたことで、「存在」「認識」と

いった抽象的な語のみならず、「使用」「説明」などの一般的な語、「結婚」「勉強」といった個人生活に関わる語、さらには「販売」「購入」「発売」といった社会・経済関連語にいたるまで、現代日本語の様々な環境を反映した汎用的な漢語サ変動詞の抽出が行われたと考えられる。

既に述べたように、無数に存在すると思われる漢語サ変動詞のうち、高頻度、かつ、汎用的に使用される語のリストはこれまで十分に整備されてこなかった。表 6 のようなデータがあれば、日本語教育において有効に活用しうるのみならず、日本語研究においても、実際の日本語で典型的に使用される項目に限った漢語サ変動詞のシンプルな用法モデルを構築することが可能になるであろう。以下、本節で得られた 93 語に限定して、漢語サ変動詞の内部構成パターンを概観していく。

4.3 RQ3 内部構成パターンの計量的概観

高頻度・汎用的漢語サ変動詞の 93 語に対して、構成要素間結合関係パターン及び品詞的結合パターンという 2 点から、その内部構成を分類した。結果として、以下の図 2~5 が得られた。

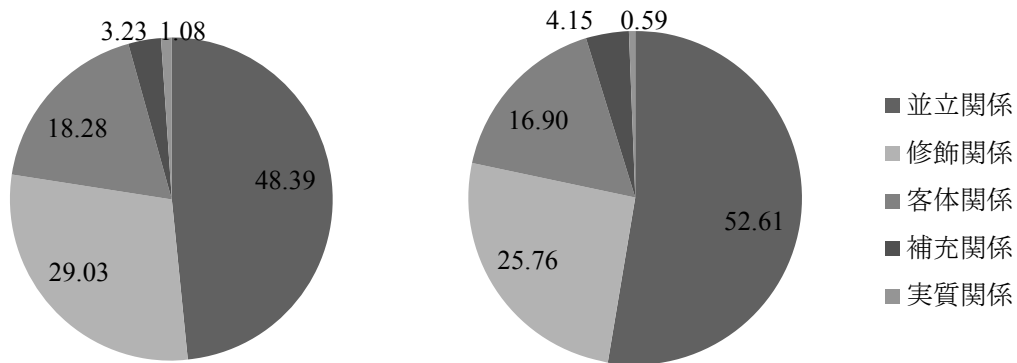


図 2・3 構成要素間結合関係パターン別比率 (左図：タイプベース；右図：トークンベース)

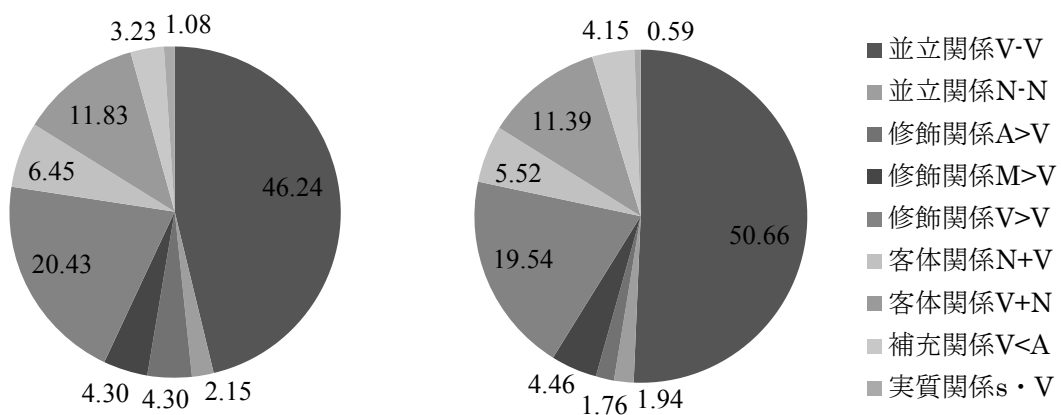


図 4・5 品詞的結合パターン別比率 (左図：タイプベース；右図：トークンベース)

先行研究は、漢語サ変動詞の内部構成について様々な分類の枠組みを提案する一方、そもそも何が代表的で典型的であるかについてはほとんど触れていなかったわけであるが、コーパスから得られた高頻度汎用的漢語サ変動詞に限定した計量分析により、当該動詞形

の内部構成の典型性について、極めて重要な知見が得られた。ここでは、4点に限って言及する。

1点目は漢語サ変動詞の内部構成の安定性についてである。タイプベースとトークンベースを比較しても、全体の傾向は大きく変化していない。このことは、漢語サ変動詞の内部構成パターンが一定の安定性を持つことを示唆する。

2点目は漢語サ変動詞の内部構成の多様性の低さについてである。各種のパタンのうち、高頻度に使用されているものはごくわずかしかない。構成要素間結合関係パターン(図2~3)について言えば、並立関係のみで全体の過半を占め、これに修飾関係、客体関係の2種を加えれば全体の9割以上を占める。また、品詞的結合パターン(図4~5)について言えば、並立関係の「V-V」だけで全体の過半を占め、修飾関係の「V>V」と客体関係の「V+N」を含めると全体の8割以上を占める。見かけ上の多様性と異なり、実際の言語運用において、漢語サ変動詞の内部構成パターンは極めて限定的であると言える。これは、語形成の観点から見て興味深い結果であると同時に、教育的観点から言えば、典型的な内部構成パターンを明示的に指導することの有用性を示唆するものでもある。

3点目は、漢語サ変動詞の語形成上の基本的特性についてである。漢語サ変動詞を構成する文字には、動詞的なもの、名詞的なもの、形容詞的なものなどが存在するわけであるが、実際には、一般の学習者が想像するように、これらが自由結合して漢語サ変動詞が作られるわけではない。並立関係の「V-V」と修飾関係の「V>V」で、全体の約8割が占められることに明らかなように、漢語サ変動詞の基本は動詞的要素を持つ2個の漢字同士の結合であると言える。教育学的には、このような漢語サ変動詞の語形成上の顕著な特性もまた学習者に適切に指導されることが望まれる。

最後に、4点目は、客体関係の品詞的結合パターンについてである。一般に、客体関係は日本語では「(S+) O+V」、中国語では「(S+) V+O」と表現されるわけであるが、「V+N」は「N+V」の2倍近く出現しており、客体関係を持つ漢語サ変動詞の内部構成は日本語というより中国語の語順に準拠していると思われる。こうした中日対照という観点からの分析は、今後の研究課題となりうる。

5. まとめと教育的示唆

本研究は、コーパスを用い、多様な日本語変種間に共通して高頻度に使われる汎用的漢語サ変動詞の特定及びその内部構成の特徴の解明を目指して議論を進めてきた。分析により、先行研究ではっきり示されていなかった一般性の高い汎用的漢語サ変動詞にどのようなものがあり、その典型的な内部構成パターンが何であるかが明らかにされた。

本研究は、日本語語彙研究におけるコーパスデータの重要性を改めて示すとともに、コーパスから得られた知見の教育応用の可能性についても一定の示唆を行うものとなった。

今後は、分析対象とするデータの範囲をさらに拡張し、分析の枠組みの一層の精緻化を図るとともに、研究で得られた成果を具体的な教材や教授法の開発に生かす方向を考察していきたい。

文 献

- 石川慎一郎(2012)『ベーシックコーパス日本語』ひつじ書房
大石亨(2012)「テキストのジャンルとメタファー表現のコレスポンデンス分析—関係のメタファーを例に」『日本認知言語学会論文集』12, pp.52-64.
沖森卓也、三省堂編修所(2013)『三省堂常用漢字辞典』三省堂
国立国語研究所(1962)『現代雑誌九十種の用語用字』秀英出版
国立国語研究所(1973)『電子計算機による新聞の語彙調査』秀英出版
国立国語研究所(1983)『高校教科書の語彙調査』秀英出版
小林英樹(2004)『現代日本語の漢語動名詞の研究』ひつじ書房
新潮社(2008)『新潮日本語漢字辞典』新潮社

- 藤堂明保、松本昭、竹田晃他 (2011) 『漢字源改訂第五版』学習研究社
- 野村雅昭 (1999) 「サ変動詞の構造」森田良行教授古稀記念論文集刊行会 (編) 『日本語研究と日本語教育』、pp.1-23、明治書院
- 橋本直幸 (2009) 「BCCWJ を利用した日本語教育語彙リスト作成の試み」特定領域研究日本語コーパス平成 20 年度公開ワークショップ (研究成果報告会) 予稿集、pp.183-190.
- 日向敏彦 (1985) 「漢語サ変動詞の構造」『上智大学国文学論集』18、pp.161-179.
- 松下達彦 (2011) 「日本語の学術共通語彙 (アカデミック・ワード) の抽出と妥当性の検証」『2011 年度日本語教育学会春季大会予稿集』、pp.244-249.
- 水本篤 (2010) 「第 8 章主成分分析：データの情報を圧縮する」石川慎一郎、前田忠彦、山崎誠 (2010) (編) 『言語研究のための統計入門』、pp.193-217、くろしお書店
- 湯本昭南 (1977) 「あわせ名詞の意味記述をめぐって」『東京外国語大学論集』27、pp.31-46.