

語構成論に基づいたサ変動詞の分類

乾 裕子 元吉 文男 井佐原 均
計量計画研究所 電子技術総合研究所 通信総合研究所

シソーラス構築のための基礎的分析として二字の漢字からなるサ変動詞を取り上げ、語構成に基づく意味的分類を行った。語構成を分析することにより、各語の上位概念を特定できるだけでなく、上位概念に相当する語の意味をどんな観点から制約するかといったより詳細な上位-下位関係を得ることができると考えられる。本稿では分析結果を報告し、上記仮説の正当性および本分析手法の問題点について考察する。

Categorization of Sahen verbs based on their morphological structure

Hiroko Inui+ Fumio Motoyoshi++ Hitoshi Isahara+++
+The Institute of Behavioral Sciences
++ Electrotechnical Laboratory
+++Communications Research Laboratory

Japanese "Sahen" verbs are those composed of morphemes, Chinese characters followed by "suru." This paper reports a categorization of Japanese Sahen verbs based on their morphological structure. By morphological analysis of Sahen verbs, one can explore not only superordinates of them but the way how each Sahen verb restricts the meaning of its superordinate. The results of our analysis is summarized and several issues are discussed.

1. はじめに

自然言語理解システム構築のためには、構文情報、意味情報に関する知識源が必要となる。筆者らは表層の構文的役割を記述した動詞概念階層の作成を行っている。動詞の階層構造を作成する場合には、動詞の分類の視点を定める必要がある[1]。本研究では、動詞の分類の基準として語構成論を採用することにより客観的な分類を試みた。

自然言語処理においてシソーラス（概念階層）から得られる言語情報は重要である。入手・活用の利便性や、階層的意味コードを有した分類体系の完成度から国立国語研究所の

分類語彙表[2]と角川の類語新辞典[3]が現在普及しているが、それぞれ意味体系の構築方法に相違がある。分類語彙表では、大分類として体言・用言・相言（・その他）の区分を設け、中分類で抽象関係・精神および行為・自然現象などの区分を行っている。それに対し類語新辞典では、分類語彙表の中分類にはほぼ相当する自然・人事・文化という区分を大分類とし、動詞・名詞といった異なる品詞に対する区分は行っていない。また編集方針について、分類語彙表では一語一義を基本に原義で分類しているのに対し、類語新辞典は語義別による重複分類を認めている[4]。望月らの研究[5]でも、この相違が両シソーラスの比較

結果として示されている。以上から示唆されるように分類基準自体の相違、あるいはどのような基準を選ぶのかという相違により異なるシソーラスが作成される。これは同時に、一つの木構造による意味体系からでは語の関連性の一部分しか捉えられないことを意味する。多次元シソーラスの必要性はこの点に起因する。

多次元シソーラスに関する研究として WordNet の構築がある。WordNet では、動詞の上位-下位関係はさまざまな種類の詳細化 (elaboration) であり、その詳細化に多様な次元 (dimension) がある [6]。多次元シソーラスの自動生成に関しては、既存シソーラスから半自動的に構築する研究 [7][8][9] や、コーパスから自動的に獲得する研究 [10] がある。川村らの研究 [9] は観点抽出に語構成を用いているが、以上の研究はすべて名詞に焦点を当てている。一方、動詞の多次元シソーラスを考える場合、Fellbaum [6] からも指摘するように動詞間の意味関係が名詞のそれよりも複雑であると見られるので、意味関係のより詳細かつ網羅的な分析が必要である。本研究ではサ変動詞の語構成に着目した仁田の研究 [11] に基づきサ変動詞を中心に意味関係の記述を行う。

本稿では、次節で先行研究と本稿での分類手法について述べる。3 節で分類結果を提示し、4 節で考察および問題点を述べる。

2. 語構成に基づいた分類

2. 1. サ変動詞の語構成

本研究で特にサ変動詞を取り上げたのは以下の見直しによる。

- 1) サ変動詞はサ変名詞の動詞用法といわれるものであり、一つの語が名詞と動詞の二つの機能を持っているという点で、異なる品詞を統合した多次元シソーラス

構築のための有効な分析材料となる

- 2) 語を構成する漢字から意味を得やすく、ある程度客観的な分析が可能である
- 3) 和語動詞に比べ多義性が少ない

本稿ではサ変動詞の意味関係だけに着目するので、名詞・動詞を統合したシソーラスについては今後の課題とする。サ変動詞は意味・用法において使用されている分野に対する依存性が高いとの報告もあるが [12]、本稿ではそれについての検討は行わない。しかし、シソーラス構築の上では考慮すべき問題である。

語構成から見た語の分類研究 [11][13][14] の中でも、仁田 [11] はとくにサ変動詞の語構成について詳細に論じている。「落馬する」が「馬から落ちる」と同義であることを指摘し、「馬から」という格成分を語の内的構造の中に含むと述べている。仁田はこのような内的構造が広くサ変動詞に見られることに着目し、その分類を試みている。仁田の分類は、格成分・状況成分・付加成分の三つに大別され、さらに以下のような下位分類を持つ。

格成分

対象・場所<着点>・場所<離点>・
場所<空間>・主体

状況成分

時の成分・所の成分

付加成分

方法・原因・期間・頻度・程度・
様態・待遇性

上記は我々の直観によく合うが、分類にあたっては小規模な調査しか行われていない。そこで、我々は仁田の分類の検証としてある程度網羅的な調査を行った。

この調査では「落馬する」のような動詞に対して「落ちる」にあたる語を上位概念動詞、「馬から」にあたる成分を制約条件と呼ぶ。また、上位概念動詞を表す語基¹を上位概念語基、制約条件を表す語基を制約語基と呼ぶ。

¹ 語を構成する漢字

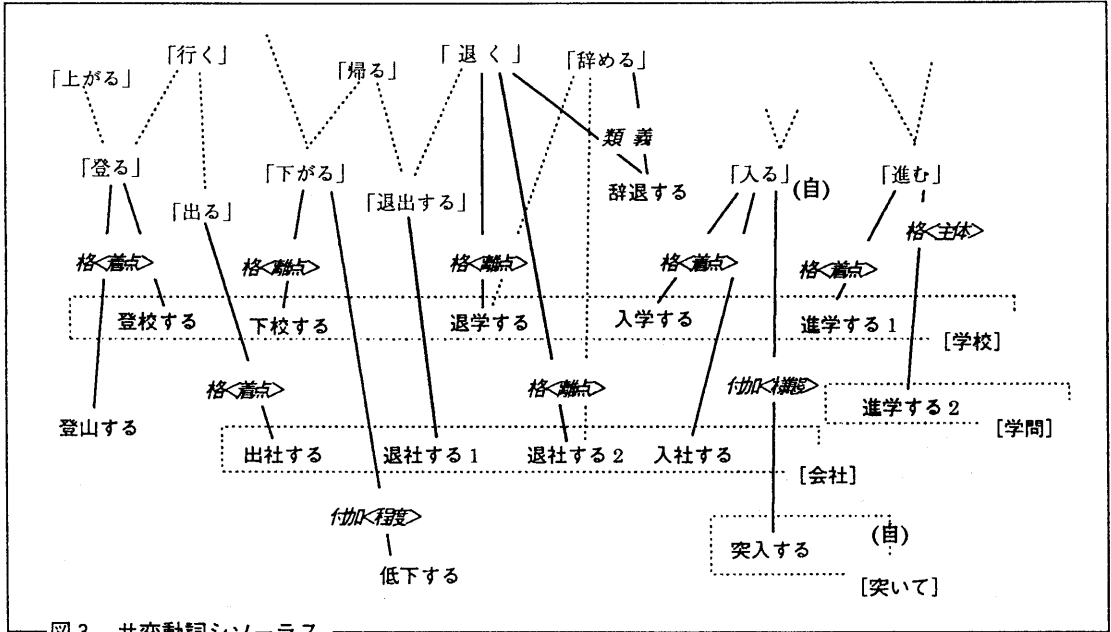


図3 サ変動詞シソーラス

- 付加<頻度>: 多発する 再演する
- 付加<程度>: 熟読する 重視する
- 付加<目的>: 出撃する 防除する
- 付加<様態>: 競演する 具現する
- 付加<待遇>: 参入する 行幸する
(分類終了)

また①の判定については語基(下線で例示)が

- a) 常用漢字かつ動詞である (ex. 「入学する」→入る)
 - b) 常用漢字ではないが動詞として使用できる (ex. 「検査する」→検(しら)べる)
 - c) 漢字を補って二字で動詞になる (ex. 「受理する」→処理する)
 - d) 一字の意味を国語辞典[15]で参照して、該当する意味があるかどうかを基準にする。(ex. 「処理する」→おさめる(理))
- c) により図4のような上位-下位関係も記述される。

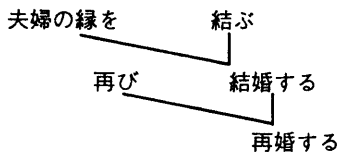


図4 「再婚する」の上位概念

3. 分類結果

本節では分類結果を統計的に要約し、また結果として得られた分類と多次元シソーラスとの関係について述べる。

3. 1. 分類結果

統計的な分類結果を表1に示す。対象とした1387語のうち、分類できなかった未分類の122語を除く1241語が現時点での有効データである。未分類の語には、1) 上位概念語基が取り出せず、かつ語基間に<類義><対義>の関係が認められないもの、2) 上位概念語基は取り出せるが、語基間の関係が分類困難であるもの、がある。未分類の語を以下に例示する。

- 例) 1) 案内する、工夫する
 2) 感謝する (cf. 謝: お礼を言う)
 提起する (cf. 提: 手に提げる)

表1 語構成によるサ変動詞分類結果

(見出し語数)

格成分		計	323
結果	10	着点	48
主体	49	難点	13
対象	186	状態	17
状況要素		計	25
時間	10	場所	15
付加要素		計	913
期間	3	方法	22
原因	5	目的	7
程度	51	様態	796
頻度	25	待遇	4
対義			20
類義			131
合計			1412語

サ変動詞 1241 語のうち自動詞・他動詞いずれの機能も持つ語 (139 語) に対しては、自動詞・他動詞に対してそれぞれの上位概念および制約条件を記述した。

例) 開会する [会] ガ はじまる (開)
[会] ヲ はじめる (開)

また一部の多義語に対してはそれぞれの意味に対する記述を行った。

例) 横行する [勝手気ままに] 行く
[勝手気ままに] 行う

これらの結果、関係記述は延べ 1412 語になった。得られた制約条件は 753 種類、上位概念の異なり 631 語、延べ 1401 語である。そのうち先に挙げた上位概念語基の判断基準 a (常用漢字かつ動詞である) を満たすものは 474 語、b (常用漢字ではないが動詞として使用できる) は 48 語、c (漢字を補って二字で動詞になる) は 39 語、d (一字の意味を国語辞典[15]で参照して、該当する意味がある) は 70 語であり、上位一下位関係の深さは最大 3 である。作業仮説に該当しない<対義><類義>の関係を除くと、サ変動詞は上位概念語基と制約語基から構成されるという作業仮説はおよそ 89% 満たされている。ただし、4 節で述べるように<様態>に分類されたものの中には制約語基を修飾成分と見

なすことが困難なものも少なくない。これらを除いた場合、作業仮説を満たすのは全体の 43% になる。

上位概念語基の判断基準 a を満たす語基は 90% である。残る 10% については上位概念の動詞抽出になんらかの無理がある。しかし、上位概念語基が存在するという仮説については、この 10% に未分類の 122 語を加えてもおよそ 80% のサ変動詞で成り立つ。

3. 2. 観点の共通性から得られる 多次元シソーラス

図 3 に示すように「入学する」「登校する」は、その上位概念語基から「入る」「登る」といった上位概念動詞を取り出すことができる。また、これらの語はいずれも「学」または「校」という制約語基を持つことから「学校」という共通の観点のもとに一つの意味クラスを作る。すなわち、いわば「学校を基準とした人の移動」という概念の下位概念動詞であると見なすことができる (図 5)。

	制約条件	上位概念動詞
登校する	[学校] ニ	登る
下校する	[学校] カラ	下がる
入学する	[学校] ニ	入る
退学する	[学校] カラ	退く
進学する	[学校] ニ	進む

図 5 共通の観点による関連語

これは「飛行機」「鳥」が「人工物」「動物」という上位一下位関係で見ると遠い関係に位置づけられるのに対し、「飛ぶ」という観点を導入することにより共通の意味クラスを作るという名詞の多次元シソーラス[8][9]と対応する。このようにサ変動詞の場合、語構成から得られる制約条件の共通性に基づいて多次元的な分類を行うことができる。

4. 考察および問題点

本節では問題点と課題について考察する。

1) 上位概念の抽出

先に示した抽出方法では「下校する」から「下がる」という上位概念は抽出できるが、「帰る」という上位概念は取れない。「帰る」を上位概念として取り出すためには「下がる」に「帰る」の意味があることを新たに参照しなければならない。今回の作業では、A BスルのA、Bそれぞれを国語辞典[15]で参照しても得られない上位概念（例：下→帰る）については記述せず、その扱いは今後の検討課題とする。

2) 制約語基の多義性

制約語基から制約条件への書き換えについて、語基間に成り立つ関係が格関係の場合には「学」や「校」から「学校」への書き換え同様必要に応じて語基を補足、または国語辞典を参照する。語基間の関係が修飾関係、特に「様態」の場合は国語辞典を参照のうえ記述する。同一の制約語基（「公」）でも以下のように複数の制約条件に対応する可能性がある。

公認する [統治機関、役所] ガ
奉公する [統治機関、朝廷] ニ
[主君] ニ
公演する 公開する 公表する 公募する
[世間] ニ
公転する [通じて用いられる、あまねく
あてはまる]

同時に上位概念語基にも同様の現象が見られる。「退社する」は「会社から帰る」「会社を辞める」の意味を持つ多義語である。上位概念動詞を「退出する」と考えると前者の語義が得られ（図3「退社する1」）、「退

職する、引退する」²と考えると後者の語義が得られる（図3「退社する2」）。今回の作業では同一の語基が多義である場合、語基の選択は作業者の判断に拠った。この判断に恣意性が含まれてしまうことは否定できない。これについても今後の検討を要する。

3) <類義>と<様態>の関係

語基A、Bの意味が国語辞典で1) 共通している、あるいは2) 類似していてサ変動詞への意味の継承に程度の差が認められない、という場合には「類義」の関係に分類している。この分類項目は仁田[11]になく、作業中にできた項目である。現在「類義」の関係には、

- 1) 更迭する： 更わる
 かわる (迭)
 検査する： 検べる [しらべる]
 査べる [しらべる]
2) 解析する： わける (解、部分部分に)
 切り離す (折、ばらばらに)
 学習する： 学ぶ
 習う

など106語が分類されている。

現時点では、「類義」という関係ではどういふ点において類似しているのかという明示的な定義がない。そのため、「様態」との境界がはっきりしないという問題がある。例えば判断に迷う例として「棄却する」「把握する」「建設する」「建築する」「増加する」「贈与する」「委任する」などがあるが、これらは「様態」に分類している。先に挙げた「類義」の各例と比較すると、それぞれの制約語基と上位概念語基との間に「時間的推移」が認められる。今回の調査では「時間的推移」のあるものを「様態」に分類している。

² 意味の違いを示すために上記の語で説明したが、恣意性についての検討が十分でないため、図3では「退く」を上位概念としている。

4) <対義>の関係

<類義>の関係同様<対義>は仁田[11]の分類にない。本稿では以下を<対義>の関係と見なした。

例) 左右する、前後する、勝負する
<対義>の関係にあるサ変動詞には上位概念動詞が存在しない。上記の例の「勝負する」では「勝」「負」を上位概念語基とみなさない。「勝つ」「負ける」は「勝負する」に含有される概念ではあるが、上位概念動詞とは認められないからである。<対義>の関係は各語基が結びついて新たな意味をなす点で派生・転義を考える際に重要である。

5) <様態>に分類される語

仁田[11]は付加<様態>に多種多様な例が分類されてしまうことを指摘しているが、本研究でも、分類困難な語が一部分類されている。

現在これらの分析を進めているが、<様態>の中には大きく分けて二つの異なる修飾成分がある。1) 動作動詞を中心とした修飾成分と、2) 様態副詞・形容詞・状態性動詞を中心としたものである。

1) 見学する 募集する (479語)

2) 予見する 静観する 狂奔する
(255語)

前者には先に述べた<類義>の関係と判別しにくいものがあり、後者には程度関係と区別しにくいものがある。

前者の仮説を満足しない要因の多くが<様態>の関係、特に制約語基が動作性の動詞に相当する場合である。これは、制約語基が担う動詞に格や共起語の情報がないので上位概念語基との関係を特定するのが難しいためと推測できる。我々は、動詞ソーラスにおける意味関係を捉える観点として統語情報を有用なものとする。IPALサ変動詞辞書[16]でもサ変動詞50語について格構造記述を試みているが、これらの詳細な統語情報の記述と

意味関係はどのように関連するかという分析が必要である。格パタン[17]・共起関係[18]など統語情報に着目して動詞のクラスタリングを行う研究も進められている。統語構造と意味の対応関係を明らかにするためにはこれらのクラスタリング結果も分析する必要がある。

今後はさらに詳細な分析が必要である。

6) サ変動詞と和語動詞の関係

<様態>の関係のように制約語基と上位概念語基がともに動詞に相当するとき、そのサ変動詞は和語動詞により近いと考えられる。通常、サ変動詞は和語動詞より多義性が少ないといわれるが、その要因は「落馬する」のように格成分を内在する点にある。語基がともに動詞に相当する場合は格成分が内在しないので、制約条件の制限がゆるい。このようなサ変動詞は語基が表す和語動詞に近い振る舞いをすると予測される。したがって<様態>の関係にあるサ変動詞の検討は和語動詞の分類基準を考える際に有効であると考えられる。

7) 動詞ソーラス構築に向けて

語構成に着目して意味を捉えることは原義主義に近い。ある語における原義から派生・転義を捉える際にも上記同様の基準軸が必要であると考えられる。異なる動詞における意味関係の基準軸と、ある一つの動詞における派生・転義の基準軸が同じ枠組みで定義できるかどうかは検討が必要であろう。これらが明らかになれば、どのような意味関係に着目した意味記述かを明示する枠組みが作成できると考える。

5. おわりに

語構成論に基づきサ変動詞を分類した。分類作業・結果・考察を経て、動詞の意味関係

を明示的に捉える際の問題点が明らかになってきた。今後はこれらの問題点を踏まえ、語の関係を捉える基準軸の作成を進める。

参考文献

- [1] 井佐原、荻野、石崎,「計算機処理のための動詞概念の分類」,情報処理学会全国大会論文集, 1987
- [2] 国立国語研究所,「分類語彙表」,秀英出版,1965
- [3] 大野、浜西,「類語新辞典」,角川書店,1981
- [4] 木村睦子,「意味分類体辞書の系譜」,「日本語学」,No12-5, 1993
- [5] 望月、本田、奥村,「語彙結束性に基づいた語義曖昧性解消の観点から見たシソーラスの比較」,情報処理学会全国大会論文集, pp3-49, 1995
- [6] Christiane Fellbaum, English Verbs as a Semantic Net, Five Papers on WordNet™, CSL Report43, 1990
- [7] 片桐、宮崎,「既存シソーラスを利用した多次元シソーラスの半自動生成法」,情報処理学会全国大会論文集, 1995
- [8] 川村、宮崎,「語を種々の観点から分類した多次元シソーラス」,情報処理学会全国大会論文集, 1994
- [9] 川村、宮崎,「既存シソーラスを利用した漢字シソーラスの半自動生成法」,電子情報通信学会研究会資料, NLC93-59, 1993
- [10] 本田、奥村,「コーパスから二重確率的シソーラスを自動的に獲得する手法の提案」,情報処理学会自然言語処理研究会資料,NL109-5, 1995
- [11] 仁田義雄,「語構成と文法記述」-漢語動詞をめぐって-,「語彙論的統語論」,明治書院, 1980
- [12] 須田、小島,「サ変動詞をめぐる2、3の考察」IPA技術報告, 1986
- [13] 森岡健二,「形態素論-語基の分類」,「上智大学国文学科紀要」1号,1984
- [14] 野村雅昭,「複合漢語の構造」,「朝倉日本語新講座1」,朝倉書店, 1987
- [15] 岩淵、西尾、水谷,「岩波国語辞典第5版」,岩波書店, 1995
- [16] 情報処理振興事業協会 (IPA) 技術センタ

ー「ソフトウェア文書のための日本語処理の研究-12」サ変動詞辞書作成手引き・サ変動詞辞書記述例, 1988

- [17] 大石 亨,「格ボタン分析に基づく動詞の語彙知識獲得に関する研究」,奈良先端科学技術大学院大学情報科学研究科情報処理学専攻修士論文, 1995
- [18] 平田、松本,「共起情報を用いた多義動詞の類別と名詞のクラスタリング」,言語処理学会年次大会発表論文集, 1995

謝辞

本稿をまとめるにあたり、情報処理振興事業協会 (IPA) 技術センターの「計算機用動詞意味記述辞書」(内部資料)と「計算機用日本語名詞辞書」を参照させていただきましたことに感謝いたします。また、貴重なコメントをくださった IPA 研究員 橋本三奈子さん、桑畑和佳子さんに感謝いたします。本研究の作業を手伝ってくださった東京女子大学の丸元聡子さん、伊藤真美子さん、遠藤千香子さんに感謝いたします。