



Title	句の包摂の用法に関する覚え書き : 検索エンジンを利用したデータ収集
Author(s)	今井, 忍
Citation	大阪外国語大学論集. 2003, 29, p. 61-70
Version Type	VoR
URL	https://hdl.handle.net/11094/79917
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

句の包摂の用法に関する覚え書き － 検索エンジンを利用したデータ収集 －

今 井 忍

On the Use of a Type of Phrasal Compound in Japanese *Gathering Data from the Internet*

IMAI Shinobu

This study is concerned with the problem of a distinction between lexical form and grammatical form. Some scholars think that they form a continuum and that there is a general tendency for a lexical form to change to a grammatical form. This being correct, it is important to describe uses of each item and to observe the whole system from a dynamic point of view, instead of native speakers' intuition. In this paper, we make use of NTT word frequency data and internet search engine in order to describe a degree to which 26 most frequently used Japanese verbs appear in a type of phrasal compound. It turns out that the more compounds are there using some verb, the more probable it will be that compounds include a phrase. This fact suggests that there is a relation between word frequency and grammaticalization. Furthermore, we make clear such facts that compounds of the same form are in fact divided into two types and the one type does not form a phrasal compound; that many compounds are technical terms of such areas as clothing, fishing, machine operation and music. Finally, some problems are pointed out concerning the construal of the relation between elements in the compound; idiomatic expressions; technical difficulty in using search engine; reliability of data on the Internet.

0. はじめに

いわゆる語彙的形態素と文法的形態素の間に截然とした区別を行うのは必ずしも容易なことではない。阪倉（1966）は、接尾語と補助動詞の関係について、「両者は、本質的に、一つづきであるといつてよいであらう。」(p.138)としている。同様のことは、関（1977）、寺村（1984）などでも示唆されている。また、通時的観点からかつて語彙的形態素であったものが新たに文法的形態素へと変化するという事実についても、阪倉（1966）に詳細な記述がある。また、近年注目されている文法化（grammaticalization, grammaticization）の研究もまた、このような観点のものであるといえよう。

このような見方に従うならば、個々の形式が実際の言語使用の中でどのように扱われているかを記述し、それらを全体として動的な視点から観察することが重要になってくる。

しかしながら、従来よくとられる、母語話者の内省ないし直感によって容認度の判断を行うという方法は、このような目的のためには適切さを欠く場合もある。なぜならば、個々の話者の判断は上で述べたような動的な変化のごく一部分に過ぎず、言語全体を反映しているとは言い難いからである。

このような点を考慮すると、大規模なコーパスを使用し、言語全体の動態を推測するという方法が有益であるように思われる。すなわち、語彙的形態素と文法的形態素の区別を反映すると考えられる現象について、個々の形式がどのような振る舞いを見せるかを観察するのである。これと類似した方法は、阪倉（1966）において、古典語の形式について適用されているが、近年のコーパスデータの充実に伴い、現代語ではより広範で網羅的な分析が可能になっている。

本稿は、コーパスの利用を試験的に行い、上記のような方法の有効性と問題点を探るものである。

1. 句の包摂について

本稿で問題にするのは、現代日本語における句の包摂という現象である。この現象についての詳細はすでに今井（1998）で述べているので、ここでは本稿で問題になる点のみを略述する。

一般に、複合語の内部には統語的な要素が介在しないことが知られている。例えば、いわゆる語彙的複合動詞の前項は受身形や使役形になることはない（「*持たれ上げる（>持ち上げる）」「*寝させ入る（>寝入る）」）。名詞句、動詞句、形容詞句などの要素についても、それらは統語的な要素であり、複合語の中には通常生じないとされる（「*[諸国漫遊の旅] だつ」、「*[子供の名] づける」、「*[駅に近] 道」）。しかし、以下のように複合語の中に句が生じているように見える場合があり、影山（1993）は、それを「句の包摂」と呼んでいる。

- (1) a. [父の墓] 参り
- b. [着物のしみ] 抜き

このような現象は、生成文法のようなレキシコンと統語論を厳密に区別する言語理論ではとりわけ問題となるものであり、理論内での解決法について様々な提案がされている（cf. Spencer 1991）。

本稿は、この現象についての理論的解決を目指すものではないため、現象の詳細には立ち入らないが、一つの点については確認する必要がある。(1)のような現象は、どのような複合語についても、また、どのようなタイプの句についても起こるものではなく、一定の制約が存在する。例えば、「墓参り」と「法案づくり」を比較すると、

- (2) a. [父の墓] 参り

- b. * [大理石の墓] 参り
- (3) a. [独法化の法案] づくり
- b. [新しい法案] づくり
- c. [今国会に提出する法案] づくり
- d. [シートベルト着用を義務づける法案] づくり

のように、前者がある一定の句しか包摂しないのに対して、後者は多くの種類の句を包摂することができる。

この現象を、0 節で述べたような観点から見ると、前者「墓参り」の後項要素である「参り」は語彙的な要素であり、後者「法案づくり」の後項要素である「づくり」はより統語的な要素へと近づいていると見ることもできるだろう。実際、阪倉(1966)は、「～めく」「～だつ」「～がる」などの要素について、多種類の語基に接しうるものほどより接尾語的であり、特定の語基にしか接しないものは複合語後項的であるとしている (p. 134)。

以下の節では、個々の句の包摂にどの程度の制約があるかを観察していくが、本稿では、今井 (1998) と同様に、〈名詞+動詞連用形〉の形式を持つ複合語のうち、「～をする」という構文をとりうるもののみを扱うことにする。それは、その他のタイプの句の包摂が基本的に接辞的なものに限られていることが多いのに対し (「～風」「～っぱい」「～めく」など)、問題の形式は形式が同一であっても包摂の程度が様々であるため、条件を比較することが容易であることによっている。

2. データの検索

データは以下のような手順で抽出した。

- (4) 『日本語の語彙特性』付録 CD-ROM より、使用頻度の高い動詞約 1000 を選び出した。
- (5) その中からサ変動詞、受身形、可能動詞、および「ついてる」「知ってる」を除外した。
- (6) 残った動詞の中で使用頻度の高いもの 26 について、
 - a. 該当動詞の連用形を後項、名詞を前項とする複合語となり、かつ、「～をする」という構文で使用可能かどうかを決定した。
 - b. a に該当するもののうち、前項名詞について包摂が起こるかどうかを決定した。

(6) の決定については、主に筆者の内省に頼ったが、実例の有無を確認するために、インターネット検索エンジン「Yahoo! Japan」のキーワード検索を利用した (検索エンジン利用の問題点については後述する)。キーワードとして、「〈動詞連用形〉をする」および「〈動詞連用形〉をして」を設定した。検索結果の中から、動詞連用形の前に名詞が隣接して複合語を成しているものを抽出し、その数を数えた。検索結果が 100 を超える場合には、100 以上の結果は無視し、100 の結果の中で抽出を行った。複合語を成すもののうち、包摂が起きていると判断できるもの (境界的と判断したものも含まれる) の実例を収集し、その数を数えた。

まず、(4) と (5) の手順で抽出された 26 の動詞のリストを使用頻度順に挙げる。

(7) する, なる, ある, いう, 出る, いる, 思う, よる, つく, 見る, みる, 受ける, 求める, 考える, 開く, 示す, 入る, 言う, 出す, 持つ, 決める, 述べる, 続く, かける, 行く

この中で「見る」と「みる」については同一のものと見なし, 「見をする」および「見をして」の形式で検索した。また, 「いる」については, 検索の都合上「居る」という漢字表記の形を使用した。

3. 結果

まず, それぞれの動詞について, 複合語の数と包摂が起きている例の数を複合語の数の多い順に挙げる。いずれも異なり度数である。

動詞	複合語の数	包摂の数
持つ	65	10
出す	53	8
開く	50	2
行く	41	1
決める	36	11
かける	30	9
受ける	14	3
見る	11	3
つく	10	2
出る	5	1
続く	4	0
なる	2	0

動詞	複合語の数	包摂の数
いう	1	0
いる	1	0
求める	1	0
考える	1	0
示す	1	0
述べる	1	0
する	0	0
ある	0	0
思う	0	0
よる	0	0
入る	0	0
言う	0	0

複合語の数と包摂の数は全体的には比例関係にあるが, 「開く」「行く」は複合語の数の割には包摂の数は少ないことが読みとれる。また, それに比して「決める」「かける」は複合語の数に対して包摂の数が多くなっている。

次に, それぞれの複合語と包摂の実例を挙げる。

動詞	複合語の数	包摂の数
持つ	灯籠持ち、鉛筆持ち、箱持ち、道具持ち、ラケット持ち、苗持ち、カバン持ち、五、六人持ち、ミット持ち、お席持ち、ベール持ち、船持ち、ちょうちん持ち、荷物持ち、タイコ持ち、ガンプレード持ち、ブラカード持ち、ケツ持ち、バス持ち、ブランデー持ち、鞍持ち、両手持ち、部屋持ち、2個持ち、ミット持ち、レフ持ち、足持ち、札持ち、ハンドガン持ち、2体持ち、ハトヤ持ち、ペンシル持ち、ボード持ち、盆持ち、猫持ち、テープ持ち、2台持ち、マイク持ち、単品持ち、多台持ち、「カゴ持ち」、ロープ持ち、ハンドル持ち、尻尾持ち、口持ち、ランプ持ち、メット持ち、天秤持ち、シングルパドル持ち、カメラバック持ち、4本持ち、サイト持ち、網持ち、片手持ち、ブツ持ち、ワイン持ち、三脚持ち、プレート持ち、衣装持ち、HP 持ち、エックス持ち、個人持ち、キンヤ持ち、中持ち、左翼持ち、	行列の道具持ち、花嫁さんのベール持ち、信助の荷物持ち、最後尾の札持ち、隊列のケツ持ち、ゴールのテープ持ち、衣装の尻尾持ち、3つのサイト持ち、8 kg オーバー 2 本のブツ持ち、予定や話しの中持ち、
出す	へそ出し、色出し、頭出し、砂出し、ごみ出し、水平出し、虫出し、芯出し、ボール出し、苗出し、形出し、お茶出し、中身出し、灰出し、つや出し、内定出し、小屋出し、品出し、鳩出し、原点出し、毒出し、荷出し、アイデア出し、炭出し、墨出し、位置出し、竿出し、根出し、型出し、商品出し、香り出し、泥出し、声出し、端子出し、ピーク出し、ダメ出し、キュー出し、指示出し、店出し、条件出し、新刊出し、バランス出し、巾出し、ネタ出し、あく出し、牙出し、形状出し、サンプル出し、雑誌出し、2枚出し、しわ出し、機能出し、空出し	切り替えミラーの原点出し、日頃溜まった毒出し、目標物の位置出し、ロッキングカーブのピーク出し、適した条件出し、大まかな形状出し、回転軸の芯出し、細かなしわ出し
開く	手開き、事務所開き、両手・両足開き、プール開き、窯開き、コース開き、相撲場開き、授業開き、神宮門開き、滝開き、グラウンド開き、鵜飼開き、街開き、土俵開き、弁当開き、道場開き、テント開き、大股開き、コタツ開き、部屋開き、鏡開き、まんじゅう開き、くす玉開き、大口開き、学年開き、茶室開き、墓開き、村開き、スキー場開き、学級開き、傘開き、両手開き、型開き、店開き、旗開き、スキー開き、キャンプ開き、岩戸開き、お重開き、カエル開き、畑開き、扉開き、括弧開き、小屋開き、床開き、樽開き、股開き、アトリエ開き、渦開き、鍋開き	役場前に構えた後援会事務所開き、後援会の事務所開き

行く	海外行き、ごみ箱行き、東京行き、小田原行き、米沢行き、龍宮行き、カリブ行き、墓場行き、新宿行き、バリ島行き、病院行き、鹿児島行き、群馬行き、浜行き、ヨーロッパ行き、九州行き、衝動行き、草稿箱行き、同地行き、GH 行き、秋葉原行き、沢行き、シカゴ行き、小浜行き、チベット行き、福岡行き、グラス行き、神島行き、撮影行き、山行き、大島行き、メトロポリタン行き、ロンドン行き、ハンブルク行き、イタリア行き、大学行き、ピックサイト行き、札幌行き、霊山観音・清水寺行き、こっち行き、大阪行き、	ヒューストンの病院行き、
決める	順番決め、位置決め、幅決め、順位決め、値決め、フレーム決め、音決め、場所決め、係決め、座位決め、場決め、型決め、座決め、アングル決め、上死点決め、レーン決め、色決め、予算決め、テーマ、グループ決め、先行決め、テーマ決め、曲決め、席決め、味決め、チーム決め、班決め、ペア決め、選手決め、卓決め、音楽決め、番号決め、仕様決め、日程決め、砂決め、長さ決め、担当決め、	磁気ヘッドの精密位置決め、釣りリンクの位置決め、集客の型決め、USB カメラ 2 台のアングル決め、子供さんのレーン決め、外装の色決め、色んな色決め、釣り座の場所決め、明日の音楽決め、ケーブルの番号決め、いろいろな仕様決め
かける	釉かけ、願かけ、声かけ、鍵かけ、ビールかけ、袋かけ、言葉かけ、砂かけ、湯かけ、針金かけ、アイロンかけ、水かけ、あみかけ、粉かけ、糸かけ、ペーパーかけ、オシッコかけ、謎かけ、シーツかけ、掃除機かけ、ワックスかけ、袋かけ、あんかけ、モップかけ、傘かけ、玉かけ、へらかけ、電話かけ、マルチかけ、巣箱かけ、	「・・・」などの言葉かけ、レッズ昇格の願かけ、「・・・」の声かけ、適当な言葉かけ、暖かい声かけ、約 200～400 枚の袋かけ、茄子の水かけ、優しい言葉かけ、+ の言葉かけ、
受ける	車検受け、上段揚げ受け、あげ受け、指示受け、注文受け、検査受け、仕事受け、ハミ受け、質問受け、答え受け、火受け、手刀受け、球受け、電話受け、	客の注文受け、野手からの球受け、わからない問題の質問受け
見る	(お) 花見、(お) 月見、模様見、様子見、ひより見、味見、よそ見、線見、柩見、カメ見、雨見	桜のお花見、桃のお花見、今後の様子見
つく	もちつき、はねつき、まりつき、たまつき、きねつき、かねつき、ずつき、ボールつき、うそつき、みそつき	お正月のもちつき、ノートルダム寺院のかねつき
出る	楽屋出、家出、旅出、磯出、村出	新しい旅出

続く	手続き、登記続き、入会続き、歌詞創り続き	—
なる	併せ馬なり、法人なり	—
いう	ものいい	—
いる	一人居	—
求める	餌求め	—
考える	素人考え	—
示す	型示し	—
述べる	星述べ	—

これらの中で、表記が異なっても同一の要素と見なされるものは一つにまとめた。例えば、「ごみ出し」には「ゴミ出し」というカタカナ表記の用例もあったが、ここでは、「ごみ出し」の中に含めている。また、「花見」と「お花見」、「月見」と「お月見」は一つの項目としてまとめた。さらに、「後援会事務所開き」のように前項名詞自体が複合語になっている場合もあったが、これは「事務所開き」と同じ扱いとし、別の複合語としては扱わなかった。

4. 考察

本稿で扱った動詞の数はわずか26であるため、問題にしている現象全体の様相を明らかにするには到底至らなかったが、前節でも述べたように複合語の数と包摂の数にある程度の比例関係があることがわかった。このことは、異なり度数で見た場合の使用頻度と文法化との間に相関があることを示唆している。

複合語の数が最も多かった動詞は「持つ」で、65の複合語が見つかったが、この中には実際には2つの異なるタイプのものが混在している。

- (8) 入場行進ではプラカード持ちをすることになっているのですが数週間前に小学校の運動会で本番に弱いことを露見していたのですごく心配でした。

(http://homepage2.nifty.com/persimmon-tree/miu/son_1/7_made/0010un.htm, 2003年6月29日)

- (9) 歯のみがきかたを習ってきました。歯ブラシは鉛筆持ちをするんだそうです。ぐーでにぎるのは、なし。(<http://www.xue.jp/diary/200205.htm>, 2003年6月29日)

(8) では、名詞と動詞が「プラカードを持つ」という関係にあるのに対し、(9) では、「鉛筆のように持つ」という関係にある。(8) と同じタイプのものとして「灯籠持ち、ラケット持ち、道具持ち」などがあり、(9) と同じタイプのものとして「ブランデー持ち、ワイン持ち、ハンドガン持ち、エックス持ち」などがある。調査データの中で包摂の例が見られたのは(8)のタイプに限られており、(9)のタイプで包摂を起こしているものはなかった。

た。

また、「行く」の例のほとんどが、「〈地名〉＋行き」という形式を持っているのも特徴的である。前節で述べたように「行く」は複合語の数の割に包摂の数が少ないが、それはこのような理由による。ここでは、異なる地名ごとに一つの複合語として数えたが、これらは基本的にまとめて一つと数えた方がよいように思われる。

複合語の中には、筆者にとって意味がよく理解できないものも含まれている。例えば、「巾出し」は、服飾関係の用語のようであるが、それが具体的にどのような行為を指しているのか筆者にとっては不明であり、したがって、「巾」と「出す」がどのような統語関係にあるのかを確信を持って言うことができない。

この例に限らず、抽出された複合語には全般に服飾、釣り、機械操作、音楽などの専門用語が多く見受けられる。このことは、複合語が名付け機能 (cf. 影山 1993) を持っていることから来していると考えられる。本稿で問題にしている複合語は、句による表現への言い換えが比較的容易であるものが多い（「花見をする」→「花を見る」, 「事務所開きをする」→「事務所を開く」, など）。それがあえて複合語で表現されているのは、それがあある特定の事象に対する名付けであるからである。このような機能が必要とされるのは、特殊な領域でのあある特定の事象であることが多いために、必然的に専門用語が多くなっていると考えられる。

5. まとめ

本稿では、句の包摂といわれる現象のごく限られた一部の領域の、さらに限られた一部のデータのみを挙げたが、この少数のデータからでも、いくつかの興味深い性質を垣間見ることができた。最後に、今後の指針のために、いくつかの問題点を挙げておきたい。

包摂の実例の中には、包摂と見るのか、あるいは、複合語全体を修飾していると見るのか判然としない例が見受けられる。例えば、「後援会の事務所開き」は、「〔後援会の事務所〕開き」と解釈すれば句を包摂していることになるが、「後援会が事務所を開く」全体が名詞句化されていると考えれば、包摂ではないことになる。「大まかな形状出し」の場合も、「大まかな形状を出す」と解釈するのか、「形状出しの仕方が大まかである」と解釈するのかによって、包摂であるかどうかが変わってくる。実際には、このような構造的に曖昧な例こそが句の包摂という特殊な現象が発現する引き金になっているとも考えられるが、この点については稿を改めて議論したい。

抽出したデータの中には、固定された表現がいくつか含まれている。例えば、「タイコ持ち」「ちょうちん持ち」「鏡開き」などである。これらは前項の意味と後項の意味から全体の意味が構成的に解釈できず（先述のように、全ての複合語が多かれ少なかれそのような性質を持っているが、これらは特にその性質が強い）、その点で他の複合語と別に扱うかどうか問題となりうる。しかしながら、山梨（1988）で指摘されている、慣用句の一部を修飾することが一般に困難であるという現象がこの場合にも当てはまると考えれば、これらは包摂を起こさないことが予測される（*「よく鳴るタイコ」持ち」「*「破れたちょうち

ん] 持ち」[*[ピカピカの鏡] 開き]。したがって、包摂という観点からは特に問題にならないと思われるが、複合語の数を問題にする場合には注意を要する。

固定表現に関連して、「カバン持ち」の例について述べておきたい。『小学館国語大辞典』によれば、「鞆持」とは「1 上役の鞆を持って供をすること。また、その人。」とある。したがって、「カバン持ちをする」という表現は「カバンを持つという行為をする」という解釈と「カバン持ちという役目をする」という解釈がありうることになる。後者の解釈はここで問題にしているものではないため、原則的に排除されなければならないが、文脈によってはこのどちらの解釈なのか決定できない場合もある。これも「タイコ持ち」と同様、複合語の数に入れるかどうかの問題となりうる。

最後に、データの抽出方法について、いくつか問題点を挙げたい。本稿では、インターネットの検索エンジンをデータ収集に用いたが、この方法には多くの問題点がある。

まず、検索がどのような条件で行われているかが利用者にとって明示的でないという点がある。例えば、2003年6月29日現在、「Yahoo! Japan」で「花見をし」と「花見をして」を検索すると、前者が3130件、後者が3920件という結果が出る。もし文字列の単純な一致によって検索しているのであれば、前者の文字列は後者の文字列に完全に含まれているため、「花見をし」の方が結果の件数が多くなるはずであるが、実際には逆になっている。これはおそらく、何らかの方法で形態素解析を施しているためと推測できるが、それがどのような性質のものであるかは明らかでない。したがって、検索結果の信頼性の根拠を問うことができなくなってしまう。このような不明瞭さは、検索結果の並び順についてもいえる。

また、今回の調査では、活用形を「する」と「して」の2つに限定したが、これは特にそうする理由はなく、より詳細なデータを得るためには全ての活用形について検索する必要があるだろう。また、ヒット件数を100で区切ったことも特に理由がなく、単に省力化のためだけである。さらに、検索する際に動詞連用形に「をする」という形式を後接させたが、それ以外の環境に重要なデータがある可能性も否定できない。このような制約を全て除くことで、網羅的なデータにより近いものが得られることになるが、逆に当該の問題に関係のないデータまでもが検索結果に混入し、それを取り除く処理に時間と労力がかかることになってしまう。

そして、これはインターネット上、特に個人が開設しているページに言えることであるが、イデオロクトや単純な書き間違いが多く含まれているために、データの信頼性が必ずしも高くないという点も問題である。今回の調査の過程でも、いくつか書き間違いと思われるデータがあったが、特に専門用語の場合には、それが正しい形式であるかどうかを判定するのは必ずしも容易でない。

このような問題点を解決するために、新聞データの利用が考えられる。今回の調査では、「CD-HIASK 朝日新聞'92」の利用を試みたが、このデータはあらかじめ抽出されたキーワードに対する一致検索であるため、本文データに対して直接文字列の一致検索ができず、今回のような調査には適していなかった。また、仮に本文データとの一致検索が可能

だったとしても、新聞というデータそのものが実際の言語使用をどの程度反映しているかには疑問が残る。例えば、前節で述べたような、専門用語が多いという一般化は新聞データからは得にくいと思われる。

したがって、より正確で網羅的なデータを、より少ない労力で得るためには、検索エンジンによる調査と新聞データの調査、および、文学作品データの調査などを複合的に行うのが現実的な選択であろう。今回の調査では、筆者の直感によるデータ作成では得られなかったであろうデータを得ることができ、また、この種の調査を行う際に問題となってくる点をいくつか発見することができた。今後の研究に役立てていきたい。

【参考文献】

- 今井忍（1998）「句の包摂に関する意味論的考察」『日本語・日本文化』大阪外国語大学留学生日本語教育センター
- 影山太郎（1993）『文法と語形成』ひつじ書房
- 阪倉篤義（1966）『語構成の研究』角川書店
- 関一雄（1977）『国語複合動詞の研究』笠間書房
- 寺村秀夫（1984）『日本語のシンタクスと意味Ⅱ』くろしお出版
- 山梨正明（1988）『比喩と理解』東京大学出版会
- Spencer, A. (1991) *Morphological Theory*. Blackwell.

(2003. 6. 30 受理)