

## コーパスを用いた論文作成のための慣用的共起表現の抽出

三國純子 小森和子

### 要 旨

本研究では、日本語母語話者が論文を書く際どのような慣用的共起表現を用いているのかを明らかにするために、日本人大学生（服装社会学及び住環境学専攻）の書いた卒業論文の抄録（計714名分、総文字数176,091字）をもとに論文コーパスを作成した。共起表現の抽出には「WATERS(Wide-ranging Automatic Text Extraction & Recognition System)」を用いた。解析の結果、本コーパスには係り受けが全部で33,087組あった。この中から「ヲ格をとる名詞と和語動詞の共起表現」2,480組を抽出し、分析を行った。その結果、「持つ」には「疑問」「関心」等、「見る」には「過程」「相違点」等、基本的な和語動詞と抽象的な名詞の共起が多いことが明らかになった。母語話者が経験的に習得しているこれらの共起表現を学習者に明示することは、学習者が日本語で論文を書く際の一助になるであろう。

キーワード：卒業論文抄録 コーパス構築 語彙教育 慣用的共起表現  
和語動詞

### 1. はじめに

国広（1997）は『理想の国語辞典』の中で、連語は従来辞書では記述の対象にされなかったが、言語使用者が勝手に作ることができない定型表現であるため、外国人学習者のためには、記述されるのが理想であると述べている。小宮（2002）も、専門分野の用語辞典には概念の記述しかないが、「公定歩合を上げる」のような連語の形にした使い方の記述が必要であると指摘している。また、大曾・滝沢（2003）は、「コロケーション」を「慣習によってまとまって使われる語の連鎖」と定義し<sup>1)</sup>、日本語学習者コーパス<sup>2)</sup>の中からコロケーションの誤用を抽出し、母語話者による大規模な日本語コーパスに依拠して、学習者に正用を提示することが望ましいと述べている。

筆者も、勤務校において、服飾やデザインを専攻する留学生の卒業論文を添削しているが、「\*やりがいを身につけた」→「やりがいを感じる」、「\*消費者意識が増えてきた」→「消費者意識が高まってきた」、「\*服は着装する人の性格、家柄見える」→「家柄を表す」等の誤用を目にすることが多い。このような誤用は、日本語学習者が日本語として定型とは言い難い共起表現を独自に作ってしまう

ことに起因する。これを防ぐには、論文で慣習的に用いられている連語形式(以下、慣用的共起表現)を提示し、指導しなければならない。それには、まず日本語母語話者が論文の中でどのような慣用的共起表現を用いているか、を調べる必要がある。しかしながら、一般に公開されているコーパスで、卒業論文レベルの学術論文を集めたものはない。また、卒業論文のような学術論文では、専攻によって用いられる学術的な専門用語は異なり、そうした専門用語を用いた慣用的共起表現も専門によって異なる。そこで、本研究では、汎用性よりも専門性を重視し、筆者の勤務校で服飾やデザインを専攻する日本人大学生が書いた卒業論文抄録集から論文コーパスを構築し、そこから、高頻度で用いられる慣用的共起表現を抽出することとした。

## 2. 語彙の誤用並びに連語に関する先行研究

佐治(1992)は、『外国人が間違えやすい日本語の研究』の中で、「\*知識を勉強する」→「知識を得る」、 「\*現代化を建設している」→「現代化に取り組んでいる」等、語の選択の誤り、動詞の慣用的表現の誤用を指摘している。佐治(1992)のこれらの誤用例は、主に中国人学習者の作文から採集したもので、手作業による語彙分析や誤用分析の結果、認められたものが収録されているが、近年では学習者コーパスを用いて誤用例を抽出した研究も行われている。

例えば、大曾・滝沢(2003)は、コロケーションによる誤用を日本語学習者コーパスから抽出し、主語と述語(「\*気分が嫌だった」→「気分が悪かった」)、ヲ格を取る名詞と動詞(「\*ビタミンを食べる」→「ビタミンをとる」、 「\*明るさをあげて」→「明るさを増して」)等の誤用を指摘している。更に、学習者コーパスを用いることで、母語話者と比較して学習者が過剰使用あるいは過少使用している語句や表現を抽出したり、学習者特有の表現を発見したり、また、母語の干渉の有無やその程度の判断等が客観的に行えるようになると述べている。また、コロケーションは「母語話者としての自然さ」に関わっており、日本語学習者も学習が進むにつれてより自然な語彙の使用が要求されるようになるが、母語話者であっても内省だけでは気づかないものもあるため、コロケーションの抽出にコーパスが力を発揮するとも述べている。

曹・仁科(2006)は神田外語大学誤用データベース及び国立国語研究所作文対訳コーパスから、形容詞・形容動詞の共起表現の誤用例 2,477 文を抽出し、「共起に関する誤用」、「語の選択に関する誤用」、「文法関係(形容動詞の接続等)の誤用」に分類した。その結果、形容詞の約 8 割(「\*能力が強い」「\*新しい顔」等)が、形容動詞の約 5 割(「\*心が親切だ」「\*経済が盛んだ」等)が、不自然な共起による誤用であった。この結果から、曹・仁科(2006)は単独の単語だけでなく、

典型的な組み合わせを提示することが有効かつ効果的だと述べている。

これらの先行研究の学習者コーパスは、誤用に基づく分析であり、それらの知見を総合すると、語彙の使用に見られる誤用の多くは不自然な共起に起因するものであると考えられ、共起表現の指導の重要性を示唆する。

ところで、こうした共起表現の指導は初級レベルだけでなく、中上級や専門教育の段階で、表現の多様化を目指す上で、一層重要であることが示唆されている。例えば、秋元(1993)は、連語を体系的・組織的に教える必要性から、『日本語表現文型』に出てくる連語 220 を 4 つに分類し、その数の割合を調べた。その結果、名詞と動詞の連語が約 80% を占めており、中でもヲ格をとる名詞と動詞の連語が多かったと述べている。また、小宮(2006)は専門分野における語彙教育として、連語を取り上げている。化学の教科書 25 冊の中から連語を抽出し、専門家が専門連語と判定したもの 1,811 種を、3 種に分類している。その結果、名詞との専門連語が 51.2%、動詞との専門連語が 42.4%、形容詞との専門連語が 6.4%であった。これらの専門語の使い方は国語辞典にも専門語辞典にも掲載されておらず、自習が困難であるため、専門連語の抽出を行ったと述べている。

以上のように、一般的な日本語学習においても、専門分野の学習においても、不自然な語彙の使用や誤用を防ぎ、より洗練された表現を習得するためには、連語の知識が必要であることが窺える。しかしながら、中級以降になると学習する語彙が多くなるため、教室活動においては、語彙の意味を確認しても、それらの語彙の使用・産出に関してまでは十分な指導、練習の時間が割けないというのが実情であろう。よって、大学で学ぶ留学生にとって、専門分野の語彙を習得し、論文を書くためには、語彙を単体で覚えるだけでなく、連語レベルでの表現の習得が必要であると言える。

### 3. コーパスの構築と解析

本研究では、留学生の卒業論文の執筆を支援することを目的に、慣用的共起表現を抽出する。抽出する元となるデータとしては、日本人大学生の卒業論文とする。そのために、まず、卒業論文コーパスを構築する。その後、構築されたコーパスから典型的な連語を抽出し、どのような語とどのぐらいの頻度で共起しているのかを調べ、慣用的共起表現を提示する。秋元(1993)で述べられているように連語の中には「ヲ格をとる動詞」が多かったことから、本研究では「名詞(ヲ格) + 和語動詞」の連語の抽出及び分析を行う。

まず、コーパス構築のために収集したデータは、大学4年生(服装社会学及び住環境学専攻)の日本語母語話者が書いた卒業論文の抄録集である。抄録は執筆者1人当たり約300字で、平成15年度239名、平成16年度238名、平成17年度237

名の計 714 名分， 総文字数 176, 091 字(3, 291 文)である。図 1 は服装社会学の抄録の例である。

環境問題がさかんに騒がれている今， ごみとして出される衣服もそのごみ問題に少なからず荷担している。その対策として循環型社会形成推進基本法やとうもろこし繊維， フリースリサイクルなどの取り組みが見られた。しかし衣服に関するリサイクルの取り組みはまだ少なく， 私達着用者の意識も低いのが現状である。このような現状をふまえた上， 本論文では環境問題に対するアパレル企業の意識や取り組みの少なさも指摘した。人間と地球が共存していくために， 衣服の廃棄を考えた上でファッションを楽しむことが今後の衣生活のあるべき姿であると主張する。

図 1 服装社会学の抄録の例

次に， コーパスの分析のために， 形態素解析を行った。形態素解析とは， 文書等を意味がわかる最小の単位(形態素)に分割し， 品詞を同定することである。日本語は語と語の間にスペースがないため， コンピュータで分析できるようにするために， まず分かち書きにしなければならないため， 形態素解析が必要となる。代表的な処理ツールとしては ChaSen (茶筌) がある。これは奈良先端科学技術大学院大学松本研究室で開発されたもので， IPA 品詞体系<sup>3)</sup>を使用している。

更に， 共起表現を抽出するには， 構文解析を行う必要がある。構文解析とは， 形態素列の相互の文法的な関係を判断する処理のことである。日本語の場合は， どの語がどの語に係っているかという係り受け関係を見出すことになる。慣用的共起表現は係り受け関係の一種であるため， 慣用的共起表現を抽出するためには， 構文解析処理が必要である。構文解析の代表的なツールとしては CaboCha<sup>4)</sup>がある。これは ChaSen をベースにした構文解析ソフトであり， 比較的効率のよい解析が行えるとされている。しかし， このソフトは新聞を対象にして解析方法を学習しているため， 領域が異なる文ではエラーが多くなる可能性があるという。

そこで， 本研究ではティ・エスコミュニケーションズ<sup>5)</sup>の構文解析ソフト WATERS (Wide-ranging Automatic Text Extraction & Recognition System) を用いた。このソフトは， 概ね IPA 品詞体系をベースにした形態素解析を行い， 更に構文解析を行う(係り受けを抽出する)ものである。WATERS は， 多くの商用ソフトに採用されており， 精度やサポート(ソフトの保守)も検証済みである。更に， 単語や係り受けを抽出するのに簡便なツールが付属されているため， 慣用的共起表現を抽出する本研究に適していると判断し， WATERS を用いることにした。表 1 は WATERS の出力例で， 太い線で囲んだ部分は係り受けを示す。

表 1 WATERS の出力例

| LeftWord | Leftpart | RightWord | Rightpart | RightSubPOS | RightTerm |
|----------|----------|-----------|-----------|-------------|-----------|
| 環境問題が    | 名詞       | 騒がれている    | 動詞        | 自立 未然形      | 騒ぐ        |
| さかんに     | 名詞       | 騒がれている    | 動詞        | 自立 未然形      | 騒ぐ        |
| 騒がれている   | 動詞       | 今,        | 名詞        | 副詞可能        | 今         |
| 今,       | 名詞       | 荷担している    | 名詞        | サ変接続        | 荷担        |
| ごみとして    | 名詞       | 出される      | 動詞        | 自立 未然形      | 出す        |
| 出される     | 動詞       | 衣服も       | 名詞        | 一般          | 衣服        |
| 衣服も      | 名詞       | 荷担している    | 名詞        | サ変接続        | 荷担        |

なお、WATERS の係り受け解析の正確さを調べるために、100 組を無作為に抽出した。その結果、7 組の係り受けが不適切であったが、そのうち 2 組は名詞と名詞の係り受けであった。例えば、「自分の 現在位置がつかみにくい場所があることがわかった (以下、下線部は WATERS が抽出した係り受けを示す)」「既製服という アパレル側の企画も変化を見せている」等である。現状のレベルの係り受け解析は、意味を無視して文法的に可能な単語の係り受けを全て同定してしまう。そのため、このように正確な判断のためには背景知識を要するような名詞と名詞の係り受けでも、文法的に可能であれば、係り受けとして判断されてしまう。また、残りの 5 組の誤分析は、名詞と動詞の係り受けであった。例えば、「暮らしを 変わらずに続けている」の「暮らしを」は、正しい係り受けの「続けている」だけでなく、「変わらずに」も抽出していた。「色彩とはどんなものかを見ていく」でも、「見ていく」という動詞は、正しい係り受けの「どんなものかを」だけでなく、「色彩とは」とも係り受け関係があると分析していた。このように、抽出された係り受けには、WATERS だけでは解析しきれない誤分析が混在していた。そこで、WATERS の分析結果を一つ一つ確認し、誤分析を取り除くという作業を行って、正しい係り受けを抽出することとした。

#### 4. 調査の結果

解析の結果、本コーパスには係り受けが全部で 33,087 組あった。その中から「ヲ格」をとる和語動詞の係り受け (連語) 2,480 組の分析結果を報告する。出現頻度の平均は 6.596、標準偏差は 18.009 であった。和語動詞の異なり語数は 376 種である。「する」動詞は漢語動詞として別途分析する必要があると考え、今回の分析には加えなかった。

石川(2006)によると、二つの語と語の結びつきの強さを測る尺度の一つとして、

MI スコア（相互情報量）が用いられる。MI スコアは、「期待値（中心語頻度×共起語頻度）」で「実測値（共起頻度×コーパス総語数）」を割り、対数に誘導した数値である。低頻度でも独特のコロケーションを特定するのに有効で、スコアが高いほど共起関係が強いとみなされる。例えば、使用頻度が 10 回前後の動詞「見せる」「遂げる」「当てる」「立てる」の共起表現の MI スコアは「焦点を当てる」（5.899）「賑わいを見せる」（5.948）「急成長を遂げる」（6.346）「仮説を立てる」（6.089）で比較的結びつきが強いと思われる。しかし、MI スコアはコーパスデータ量が少ない場合、低頻度の共起表現を過大評価してしまう危険性がある。本コーパスで最も MI スコアの高い共起表現を抽出したところ、「暖簾をはずす」10.406 であった。「暖簾」「はずす」はそれぞれ使用頻度が 1 回ずつで、共起頻度も 1 回であるため、MI スコアが高く算出されたのである。他にも、「生き残りをかける」、「礼節をわかまえる」、「はぎれを継ぎ合わせる」、等が MI スコアでは高得点となり、データ量の少ない本コーパスでは MI スコアが共起表現を示す適切な指標とは言い難い。よって、本調査では頻度を中心に分析を進めていくことにする。

出現頻度の高い和語動詞上位 20 語を見ると(表 2)、「行う」「持つ」「見る」等、初級段階で学習するものが多いことがわかる。

表 2 出現頻度の高い和語動詞（上位20種）

| 順位 | 動詞  | 日能級 | 出現頻度  | 順位 | 動詞    | 日能級 | 出現頻度 |
|----|-----|-----|-------|----|-------|-----|------|
| 1  | 行う  | 3   | 2 3 7 | 11 | 楽しむ   | 3   | 4 0  |
| 2  | 持つ  | 3   | 1 4 5 | 12 | 求める   | 2   | 3 7  |
| 3  | 見る  | 4   | 1 1 7 | 13 | 受ける   | 3   | 3 4  |
| 4  | 感じる | 2   | 9 4   | 14 | 取り入れる | 2   | 3 3  |
| 5  | 与える | 2   | 7 0   | 15 | 探る    | 2   | 3 2  |
| 6  | 考える | 3   | 5 6   | 16 | 知る    | 4   | 3 0  |
| 7  | 作る  | 4   | 5 3   | 17 | 目指す   | 2   | 2 9  |
| 8  | 用いる | 2   | 5 3   | 18 | いかす   | 1   | 2 8  |
| 9  | 得る  | 2   | 4 5   | 19 | 使う    | 4   | 2 6  |
| 10 | 設ける | 1   | 4 2   | 20 | 凶る    | 1   | 2 6  |

注: 日能級とは、日本語能力試験出題級の略

また、日本語能力試験の出題級では(表 3)、全 376 種のうち、216 種(57.29%)が 2～4 級で、1 級までを含めると、283 種(75.07%)であった。級外語彙の中には、「生み出す」「見出す」「読み取る」「兼ね備える」等、1～4 級の動詞から成る複合語が多く含まれていた。なお、複合動詞は、個々の動詞としては『日本語能力試験出題基準(改訂版)』(2002)に記載されていても、複合語の形で記載がな

ければ（「あわせ持つ」「読み解く」「くみ取る」等）、級外とした。ただし、複合動詞の個々の動詞（「鑑みる」「きたす」「極める」「ひもとく」等）で級外のもののは39種(10.37%)のみであったことから、卒業論文で使用されている和語動詞の約90%は、日本語能力試験の語彙表に含まれていることになる。

表3 日本語能力試験出題級

単位:頻度

|     | 4級        | 3級        | 2級         | 1級        | 級外        | 合計          |
|-----|-----------|-----------|------------|-----------|-----------|-------------|
| 動詞種 | 39(10.37) | 47(12.50) | 130(34.58) | 67(17.82) | 93(24.73) | 376(100.00) |

注:( )は全376種に占める割合

次に、使用頻度の高い動詞がどのような名詞と共起しているかを調べた。その結果、動詞は初級段階で学習する基本的なものが多いが、共起する名詞には抽象的なものが多いことが明らかになった。表4にその一例を示す。

表4 動詞と共起する名詞の例（一部抜粋） ( )は出現頻度

| 動詞      | 共起する名詞 |          |         |        |        |
|---------|--------|----------|---------|--------|--------|
|         | 1位     | 2位       | 3位      | 4位     | 5位     |
| 行う(237) | 調査(78) | 研究(26)   | 提案(16)  | 分析(12) | 実験(12) |
| 持つ(145) | 興味(19) | 機能(9)    | 繋がり(5)  | 意味(5)  | 関心(3)  |
| 与える(70) | 影響(34) | 印象(4)    | イメージ(3) | 変化(2)  | 感動(2)  |
| 受ける(34) | 影響(19) | 刺激(2)    | 印象(2)   | 衝撃(1)  | 感銘(1)  |
| 図る(26)  | 交流(6)  | 差別化(2)   | 活性化(2)  | 合理化(1) | 確立(1)  |
| 送る(15)  | 生活(8)  | メッセージ(2) | 生涯(1)   | 人生(1)  | 情報(1)  |
| 見せる(13) | 賑わい(3) | 身体(2)    | 自分(2)   | 変化(2)  | 広がり(1) |
| 遂げる(11) | 変化(5)  | 急成(2)    | 変貌(1)   | 発展(1)  | 発達(1)  |
| 当てる(11) | 焦点(6)  | スポット(4)  | 視点(1)   |        |        |
| 立てる(10) | 仮説(6)  | 計画(2)    | 予想(1)   | 項目(1)  |        |

各動詞について、共起頻度の高い名詞の日本語能力試験の出題級を調べると、動詞と共起名詞とで難易度に差があるものがあることがわかる。例えば、「持つ」「見せる」は4級動詞、「行う」「受ける」「立てる」は3級動詞であるが、共起するのは、それより難しい2級、1級ばかりである。例えば「持つ」は、初級では「かばんを持つ」「傘を持つ」等、＜具象物を手で掴む＞という意味で提示することが多く、それらの「かばん」、「かさ」、といった名詞も4級、3級の語彙であるが、本コーパスでは「興味」「機能」「繋がり」等、2級の抽象的な名詞と共起することが多かった。これは、「持つ」が多義的であり、＜具象物を手で掴む＞という意味のほかに、＜ある属性、態度を備える＞という意味があり、論文ではこの

意味での用法が多いためだと考えられる。「空間」や「座敷」等の具象物との共起も認められたが、場所を表す名詞が主であった。これは、本コーパスが住環境学の卒業論文であることに起因すると言えよう。

よって、こうした共起表現の場合、「興味」や「機能」といった名詞を提示した時に、それらと共起するのは、既に初級で習った4級や3級の「持つ」「行う」等の平易な語であり、それらの動詞が多義的であることを学習者に教える必要があると思われる。

また、「図る」では、接尾辞の「化」の付く名詞が多く、「送る」も物の輸送ではなく、「時を過ごす」という意味での用法が多いことがわかる。また、「与える」と「受ける」では共起する名詞が共通しているものがある。よって、この2つの動詞は同時に教えることが可能であり、また効率的であることが示唆される。

なお、本コーパスの専門分野に特徴的な表現も抽出された。例えば、「ジャポニスム現象を起こす」「賑わいを見せる」等である。これらの「起こす」「見せる」は、日本語学習の初期段階で学ぶ難易度の低い動詞である。しかし、学習当初は「子供を起こす」「教科書を見せる」等の具象名詞と共起する形で学ぶことが多いため、これらの基本的な動詞がどのような抽象的な言葉、特に漢語と共に用いられるのか、学習者が十分に習得しているとは言い難い。また、教育現場においても、既習である「起こす」「見せる」を、上級レベルになって改めて指導する機会も多くはないだろう。更に、論文でもこのような基本動詞が用いられているということを知る機会も少ないのではないかとと思われる。そのため、こうした定型表現はなかなか習得されず、学習者は文法的には正しくても日本語としては慣用的とは言い難い、不自然な共起表現を独自に作ってしまうのではないかと考える。

## 5. 他のコーパスとの比較

次に、本研究の論文コーパスで抽出された高頻度の動詞が、その他のコーパスではどのような名詞と共起して使われているかを探索する。これによって、本研究が抽出した共起表現が、本専門分野の論文に特有の共起表現であるか否かについて、比較検討できる。他のコーパスとしては、先ごろその一部がデモ版として公開された書き言葉均衡コーパス(以下、均衡コーパス)を用いた。均衡コーパスは、文部科学省の特定領域研究と国立国語研究所の「kotonoha」計画の共同研究によるものである(<http://www.kotonoha.gr.jp/demo/>にて公開)。均衡コーパスのデモ版は、政府刊行白書(以下、白書)とインターネットサイトのyahoo知恵袋(以下、yahoo)から成っている。白書は書き言葉によるコーパスであるが、yahooはブログ等を含んだインターネットサイトであるため、純然たる書き言葉とは言いがたく、話し言葉を文字化したものと捉えられる。今回の分析では、白書と



yahoo の中で、2004 年から 2005 年までの 2 年間のデータを対象とした。なお、デモ版のため、検索ヒット件数は 500 件が上限である。

分析は本論文コーパスの上位 10 の動詞について行ったが、本稿では、本論文コーパスと共通点や相違点が顕著な 6 つの動詞について考察する(表 5)。まず、本研究で最も使用頻度が高かった「行う」では、白書で頻度が高かった共起名詞は、「支援」「放送」「検討」等、漢語で比較的抽象的な概念を表す語が多かった。また、本研究の論文コーパスで上位の「研究」も抽出された。一方、yahoo では、具体的な活動を表す「運動」「行為」「手続き」「工事」等が多く、論文コーパスと重なる語はなかった。また、「持つ」では、白書も yahoo も「子供」が最も多く、論文コーパスとは異なる結果であった。ただし、いずれのコーパスを見ても、「持つ」が「具象物を手で掴む」という用法では用いられておらず、「所有する」「備える」という意味での使用が高頻度であることがわかる。

表 5 「書き言葉均衡コーパスにおける共起表現」

| 動詞  | ジャンル  | 共起する名詞                |
|-----|-------|-----------------------|
| 行う  | 論文    | 調査, 研究, 提案, 分析, 実験    |
|     | 白書    | 支援, 放送, 検討, 活動, 研究    |
|     | yahoo | 運動, 行為, 手続き, 工事, 活動   |
| 持つ  | 論文    | 興味, 機能, 繋がり, 意味, 関心   |
|     | 白書    | 子供, 技能, ニーズ, 能力       |
|     | yahoo | 子供, 興味, 自信            |
| 与える | 論文    | 影響, 印象, イメージ, 変化, 感動  |
|     | 白書    | 影響, 効果, 権限, 支持, 潤い    |
|     | yahoo | 影響, 刺激, 休暇, 愛, イメージ   |
| 図る  | 論文    | 交流, ~化, 確立            |
|     | 白書    | 充実, ~化, 交流            |
|     | yahoo | ~化, 安定, イメージアップ       |
| 送る  | 論文    | 生活, メッセージ, 生涯, 人生, 情報 |
|     | 白書    | 生活, 人生, 情報, 意見, 暮らし   |
|     | yahoo | メール, 写真, 現金, 商品, 本    |
| 見せる | 論文    | 賑わい, 身体, 自分, 変化       |
|     | 白書    | 傾向, 高まり, パフォーマンス, 格差  |
|     | yahoo | I D, ビデオ, ロゴ, 見本, 弱み  |

一方、「与える」や「図る」では、均衡コーパスと本論文コーパスに共通点が認

められた。「与える」の場合、いずれのコーパスでも最も多かったのは「影響」であった。「図る」も同様に、接尾辞「化」のつく「強化」「差別化」「多様化」や、「交流」が多かった。すなわち、「与える」や「図る」は、どのようなタイプの文章においても、共起する名詞が同じであることがわかった。

また、硬い書き言葉の共通点と見られる現象もあった。「送る」では、本論文コーパスも白書も、「生活」「人生」等の時間に関する名詞が上位で、「送る」が「過ごす」の意味で用いられていることがわかる。一方、yahoo では、具象物や「メール」と共起し、「郵送・輸送」の意味で用いられている。更に、「見せる」では、コーパス間で全く同じ語が認められなかった。但し、論文コーパスや白書では、「見せる」の中心的意味と考えられる「具象物を相手に見させる」という意味ではなく、「そのような状態を呈する」という意味で使われることが多いという点で共通性がある。一方、yahoo では「具象物を相手に見させる」という意味で用いられることが多いのがわかる。

## 6. まとめと今後の課題

今回の調査結果から、卒業論文レベルでは、2級語彙が多く占めるということがわかった。よって、まずは、2級レベルの語彙の習得を目指す必要があると考えられる。また、抽出された和語動詞は、概ね基本的かつ初歩的な動詞が多いのに対して、共起する名詞は抽象的且つ難易度の高いものが多く、動詞と名詞の組み合わせにアンバランスが生じていることがわかる。よって、教室活動においては、中級段階でこれらの抽象的な名詞が出てきた時に、共起する動詞が初級レベルの動詞であり、それらの基本動詞が様々な名詞と共起し得ることを明示することが重要であると考えられる。論文においても、こうした初級レベルの基本動詞が多用されるのは、これらの動詞が複数の意味を有する多義語であることに起因すると考えられる。よって、初級の導入時に当該語の基本義を提示した後、中級段階で、再度これらの動詞の複数の意味を提示し、共起名詞との組み合わせを指導することが有効だと考える。

また、均衡コーパスと試験的に比較したところ、論文コーパスの語彙使用の特徴の一部を垣間見ることができた。慣用的な共起表現には、文章のタイプによって異なる個別的なものと、文章のタイプによらず汎用的なものがあることがわかった。また、多義的な動詞の場合、文章のタイプによってよく用いられる意味が異なる傾向があることが示された。今後は他の専門分野や他のジャンルのコーパスとの比較を行うことによって、共起表現と文章のタイプとの関係について検討する必要があると考える。その際、動詞の語彙概念構造を分析し、共起する名詞の意味特性を言語学的に分析することも重要である。

更に、本研究では、名詞と和語動詞の係り受けを分析したが、ヲ格をとる漢語動詞やニ格をとる動詞についても、今後同様に分析を行いたい。また、より信頼性のある抽出を行うためには、コーパスデータ量を増やす必要があるだろう。母語話者の論文コーパスを分析することによって、母語話者が経験的に習得している共起表現を抽出し、学習者に明示することは、学習者が日本語で論文を書く際の一助になるであろう。

## 付記

本論文は第 16 回小出記念日本語教育研究会(2007 年 6 月 30 日、於：東京女子大学)にて発表した内容に加筆修正したものである。本研究は、平成 17～19 年度文部科学省科学研究費萌芽研究(研究代表者：三國純子、課題番号：17652055)による研究成果の一部である。また、本研究の一部は、東京大学 21 世紀 COE「心とことば—進化認知科学的展開」の助成を受けた。

## 謝辞

WATERS の使用、ならびにサポートにご尽力いただいた株式会社ティ・エスココミュニケーションズの高井貞治氏、小山由記氏に厚く御礼申し上げます。また、データ処理にご協力いただいた東京大学大学院の渡邊卓也氏にも感謝致します。

## 注

- 1) 大曾・滝沢(2003)が指摘するように、「コロケーション」の定義は研究者によって異なっている。例えば、Nation(2001:317)は、“collocation”を広く捉え、「1つのまとまりを成す単語の群」と定義している。組み合わせの語が内容語か機能語か、意味が個々の単語から引き出せるか否か等、collocationにおける二単語間の共起形式、関係性、関係強度は多様であり、厳密な定義、分類が難しいと述べている。但し、Nation(2001:317)は、言語学習という点から見れば、「共起頻度が高く、個々の単語から意味が推測しにくいもの」と捉えるのが有効だとしている。また、三好(2007:81)は、「語と語の結びつきには、強さによって3つの段階がある」としている。(1)慣用表現、(2)固定的だが個々の単語の意味から全体の意味が推測できるもの(「傘をさす」)、(3)語の意味範囲で自由に結びつくもの(「お茶を飲む」)である。これは、国広(1997:128)の、「成句」、「連語」、「用例」にそれぞれ対応する。しかし、三好(2007)は、(3)の自由連結は学習者が産出する場合に必要な知識であるため、(3)の自由連結も連語として捉えて、指導する必要があると述べている。本研究は、日本語学習者の論文執筆支援という産出の側面を念頭に入れており、三好(2007)の視点に

符合する。そこで、本研究でも、統語的に共起し得る、慣用的なまとまりを成す語と語の組み合わせ、すなわち、(2)と(3)を合わせて、“collocation/コロケーション/連語”と捉える。

- 2) コーパスとは、人間が産出した言葉や文書(テキスト)を大量に集めた言語データのことである。
- 3) IPA(Information-technology Promotion Agency)とは独立行政法人情報処理推進機構のことである。IPA 品詞体系とは、用言の活用形を含めて約 500 種類に品詞を判定する精度の高い分類基準である。日本語教育で行われている品詞分類と異なり、形容動詞は名詞の形容動詞語幹として、すなわち、名詞に分類される。
- 4) CaboCha も ChaSen 同様に奈良先端科学技術大学院大学松本研究室で開発された日本語係り受け解析システムである。  
<http://chasen.org/~taku/software/cabocha/>
- 5) 株式会社ティ・エスコミュニケーションズは、自然言語処理の技術開発を行っている会社である。WATERS は入力文から単語や、単語相互の係り受け関係を取り出すことによって、自然文を定量的に計測するデータを提供する自然文解析エンジンで、さまざまな商用ソフトに採用されている。  
<http://www.web-tscom.co.jp/html/waters.html>

## 参考文献

- 1) 秋元美晴(1993)「語彙教育における連語指導の意義について」”Proceedings of the 4<sup>th</sup> conference on second language research in Japan” 国際大学, pp. 29-47.
- 2) 石川慎一郎(2006)「言語コーパスからのコロケーション検出手法—基礎的統計値について」『統計数理研究所共同研究レポート』, 190, pp. 1-14.
- 3) 大曾美恵子(1999)『日本語学習者の作文コーパス：電子化による共有資源化』(研究課題番号 08558020)平成 8～10 年度科学研究費(基盤研究)研究成果報告書.
- 4) 大曾美恵子・滝沢直弘(2003)「コーパスによる日本語教育の研究—コロケーション及びその誤用を中心に」『日本語学』4 月臨時増刊号, pp. 234-244.
- 5) 大曾美恵子(2005)「コーパスによるコロケーションの特定—日本語学習辞書の充実を目指して—」『レキシコンフォーラム』1, pp. 11-23.
- 6) 荻野綱男(編)(2007)『コーパスを利用した国語辞典編集法の研究』平成 18 年度文部科学省科学研究費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21 世紀の日本語研究の基盤整備」報告書.

- 7) 国広哲弥(1997)『理想の国語辞典』大修館書店.
- 8) 小宮千鶴子(2002)「専門連語と専門連語辞書」『情報知識学会誌』12(1), pp. 20-31.
- 9) 小宮千鶴子(2006)「理工系留学生のための科学の専門連語—高校教科書の調査に基づく選定—」『講座日本語教育』42, pp. 154-169.
- 10) 国際交流基金・日本国際教育協会(2002)『日本語能力試験出題基準(改訂版)』凡人社.
- 11) 齋藤俊雄・中村純作・赤野一郎(1998)『コーパス言語学—基礎と実践—』研究社.
- 12) 佐治圭三(1992)『外国人が間違えやすい日本語の表現の研究』ひつじ書房.
- 13) 鈴木智美(1999)「意味的な誤用に見られる主な傾向—慣習的に定着した表現及び類似の表現に関わる誤り—」『日本語学習者の作文コーパス：電子化による共有資源化』(研究課題番号 08558020). pp. 131-145.
- 14) 曹紅荃・仁科喜久子(2006)「中国人学習者の作文誤用例から見る共起表現の習得及び教育への提言—名詞と形容詞及び形容動詞の共起表現について」『日本語教育』130, pp. 70-79.
- 15) 外池俊幸(1999)「日本語学習者の作文コーパス：コロケーションが関係する諸問題」『日本語学習者の作文コーパス：電子化による共有資源化』(研究課題番号 08558020). pp. 90-101.
- 16) 滝沢直宏(1999)「コロケーションに関わる誤用—日本語学習者の作文コーパスに見られる英語母語話者の誤用例から—」『日本語学習者の作文コーパス：電子化による共有資源化』(研究課題番号 08558020).
- 17) 松野和子・杉浦正利(2004)「コロケーションの定義—コロケーションの概念と判定基準に関する考察—」(オンライン)  
(<http://sugiura3.nagoyau.ac.jp/project/nnscollocation/report/MatsCollocation-sugi.pdf>)
- 18) 松本裕治・徳永健伸(2000)「コーパスに基づく言語処理の限界と展望」『IPSJ(情報処理学会)Magazine』41(7), pp. 793-796.
- 19) 三好裕子(2007)「連語による語彙指導の有効性の検討」『日本語教育』134, pp. 80-89.
- 20) Nation, I. S. P. (2001) *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.

(三國：文化女子大学，小森：国際交流基金)