

# 国立国語研究所学術情報リポジトリ

A comparison of Japanese compound verb usage by native speakers and learners of Japanese as a second language : A corpus-based approach

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: 陳, 曦, CHEN, Xi メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00002184">https://doi.org/10.15084/00002184</a>

# 学習者と母語話者における日本語複合動詞の使用状況の比較

—— コーパスによるアプローチ ——

陳 曦

(名古屋大学大学院生)

キーワード

複合動詞, 学習者コーパス, 話し言葉コーパス, 第二言語習得

## 要 旨

日本語学習者の話し言葉における日本語複合動詞の使用状況を調査した。本研究では、第二言語習得研究にコーパスを利用することに着目し、「学習者コーパス」と「母語話者コーパス」の二種類のコーパスを用い、それらを比較して行った。本稿の目的は、日本語複合動詞の使用状況を明らかにすることと、現状のコーパス利用の実証と考察とにある。調査の結果、日本語複合動詞は学習者全体で使用頻度も種類も少なく、特に「差し上げる」、「申し上げる」など敬語を表す用法が少ないこと、また、学習者と母語話者の使用頻度上位の複合動詞は多くが共通する一方、学習者は「～あう」、「～だす」の使用は多いが、「～はじめる」、「～つづける」などのアスペクトを表すものの使用は少ないという違いのあることなどがわかった。学習者の熟達度や学習者の母語の違いによっても使用頻度と種類について有意な差が認められた。また、現状のコーパスには表記のゆれ、属性情報の精緻化、誤用情報の付加などに課題が残ることを指摘し、今後のコーパス整備における課題を提案した。

## 1. はじめに

本研究は、日本語学習者（以下「学習者」と記す）の話し言葉における日本語複合動詞（以下「複合動詞」と記す）の使用状況を調査したものである。本研究では、その調査を「学習者コーパス」と「母語話者コーパス」を用いて両者を比較することによって行った。

本稿でいう複合動詞とは、「言い出す、思い込む」のように前の動詞（以下「前項動詞」と記す）の連用形にもう一つの動詞（以下「後項動詞」と記す）が結合したものを指す。複合動詞はその結合条件、結合した複合動詞の相互作用などが複雑で、学習者にとってその理解と使用は難しく、有効な習得支援が必要であると先行研究で指摘されている（松田 2004）。しかしながら、複合動詞を習得研究の対象として扱う研究は、これまで、寺田(2001)、松田(2002a, 2002b, 2004)、陳(2004)とまだ少なく、学習者の複合動詞全般の使用状況を実際の発話データを用いて調査したような研究はほとんどないというのが現状である。複合動詞教育や習得支援の方向性を探るためには、まずは学習者の使用状況を調査することが必要であると考えられる。

本稿の目的は大きく二つある。第一の目的は、複合動詞習得のために、その複合動詞の使用状況を明らかにすることである。そして、第二の目的は、習得研究におけるコーパス利用の実証と

考察である。

本稿では、まず、現状で習得研究に利用できる主なコーパスを紹介し、本研究におけるコーパス選定と、具体的な利用方法について述べる。次に、学習者コーパスと母語話者コーパスの比較調査から明らかになった複合動詞の使用状況について述べる。そして、その後、今回の調査を通じてわかった現状のコーパスの問題点を指摘し、今後のコーパス整備の課題についての考察を述べる。

## 2. 複合動詞習得の先行研究と現状が抱えている課題

これまで、複合動詞に関しては、結合条件や分類に関する体系的な研究、複合動詞の意味的側面に関する意味的研究、及び他言語との対照研究などが盛んに行われてきた（詳細は姫野1999、松田2004を参照）。それらに対し、習得研究はまだ緒についたばかりである。ここで寺田(2001)と松田(2004)を紹介し、複合動詞習得研究として次に取り組むべき課題について考える。

### 2.1. 複合動詞習得に関する先行研究

寺田(2001)は、影山(1993, 1996)、影山・由本(1997)の複合動詞には「統語的複合動詞<sup>1</sup>」と「語彙的複合動詞」があるという主張を受けて、二者は、異なるメカニズムによって習得されるという仮説をたて、帰国子女を対象に複合動詞の習得に関する調査を行った。「統語的複合動詞」は規則を適応していく default タイプであり、「語彙的複合動詞」は頻度と類推によって長期記憶に載せる (= mental lexicon<sup>2</sup> にリストして覚える) タイプであろうとする習得仮説を立て、帰国子女を対象に、複合動詞の作成、理解、判定テストを質問紙で調べた。その結果、予想どおり海外滞在年数の多寡に関わらず「統語的複合動詞」は誤りが少ないのに対し、「語彙的複合動詞」は海外滞在年数が長いほど誤用が増えたと報告している。

松田は、複合動詞の習得に焦点を当て、複合動詞そのものの研究と複合動詞の習得研究について一連の研究を挙げている (2002a, 2002b, 2004)。ここでは、松田の一連の研究のまとめである松田(2004)を紹介する。

松田(2004)は、日本語学、言語学における複合動詞の研究を概観した上、教育の立場に立ち、複合動詞の意味に注目し、「～こむ」について深く意味研究と習得研究を行った。具体的には、認知意味論の手法を援用し、「～こむ」の多義的語義全体を包含するような共通の図式(コア図式<sup>3</sup>)で統一的に説明しようと試みた。「～こむ」の意味用法を四種類に分け、学習者は「～こむ」についてどのような意味知識を持っているかを自らの分類を用い、理解の面を文法性判断テストで、産出の面を作文テストで調査を行った。学習者により産出された作文を母語話者に判断してもらい、その結果をもとに分析した。調査の結果から学習者の理解の面も産出の面も問題が多く、複合動詞の習得がうまくできていないことが分かった。その原因としては、学習者は母語を問わず、「～こむ」の使用の拠り所となる認知的基盤の獲得が十分になされていないと指摘した。

## 2.2. 次に取り組むべき課題

寺田も松田も従来扱われることの少なかった複合動詞の習得を扱っている点で先駆性があり、その点では十分評価できる。寺田の研究は、複合動詞の言語学における成果を踏まえ、また習得理論を踏まえた習得研究として価値がある。松田の認知意味論によるアプローチは複合動詞習得研究に大いに貢献しうると思われる。

しかし、寺田と松田を含め、これまでの先行研究が複合動詞全般を対象に、特に、(1)量的な使用状況の調査がされていないこと、(2)話し言葉を対象にした複合動詞の使用状況の調査を扱っていないという点を、次の課題に向かう問題点として指摘することができる。複合動詞をいかに教材に取り込んでいくべきか、いかに教えるべきかなどを探るために、まず学習者がどのように複合動詞を使用しているか、どのような使用傾向があるかなど、使用実態の把握が必要になると思われる。そこで、本研究では、この「複合動詞の話し言葉における使用状況に関する量的調査」を課題として取り上げる。話し言葉における複合動詞使用状況を見るためには、当然のことながら、大量の自然産出に近いデータを分析することが望ましい。そこで、「学習者コーパス」と「母語話者コーパス」という二つタイプのコーパスに着目し、それらの有効活用を試みる(コーパスについては3章に詳述)。

## 2.3. 本研究の課題

以上のことを踏まえ、本研究では、学習者コーパスと母語話者コーパスとの比較で、学習者の話し言葉における複合動詞の使用状況を明らかにすることを目的とする。具体的な研究課題は以下の五つである。

- 1) 学習者による複合動詞全体の使用頻度は母語話者と同程度か。
- 2) 母語話者と比べ、どのような後項動詞を使用する傾向があるか。
- 3) 母語話者と比べ、どのような前項動詞を使用する傾向があるか。
- 4) 学習者の熟達度による複合動詞の使用状況の違いはあるか。
- 5) 学習者の母語別による複合動詞の使用状況の違いはあるか。

また、上記に加え、学習者コーパスと母語話者コーパスという二つのコーパスを利用する上で明らかになった問題点を指摘し、今後のコーパス整備や活用の方向性についても考察する。

## 3. 日本語学習者コーパスと第二言語習得研究

本研究で用いるコーパスの概要とその選定理由を述べる前に、まず、学習者コーパスの定義、位置づけ、日本語学習者コーパスの現状及びその研究例について紹介する。

### 3.1. 学習者コーパスの定義、位置づけ

学習者コーパスとは、学習者の産出した「テキスト」を体系的に集めた電子のデータである(Nesselhauf 2004)。この場合の「テキスト」には書き言葉だけではなく、話し言葉も含まれ、具体的には、自由作文やインタビューによるデータなどが含まれる。「体系的に集めた」という

点に関して、Granger et al. (eds.) (2002)は「学習者コーパス」の構築基準として学習者（学習環境、母語、他の学習言語、熟達度）とタスク環境（時間制限、辞書などの使用、試験かどうか、聞き手は誰か）に関する情報がデータに含まれているべきであると述べている。

従来、第二言語習得研究のためのデータ収集方法として、実験による誘出法、文法性判断テスト、母語話者や研究者の直感判断に依拠する内省法などの方策が用いられてきた。それが、1990年代後半から、「自然」産出データとしての学習者コーパスと呼ばれるものが、コンピュータの普及とともに作成され、注目されてきている。

学習者コーパスは、通常、各研究者が個別に収集した発話事例とは比較にならない規模があり、かつ、すでに電子化されていることにより、直ちにコンピュータを用いて分析を行うことが可能である。また、データを産出した個々の学習者の熟達度や母語など学習者の背景に関するデータも有するため、習得研究に大いに寄与することができると考えられる。

### 3.2. 日本語学習者コーパスの現状

大曾(2006)によると、日本語教育に生かせるコーパスとして、(1)日本語母語話者コーパス(2)日本語学習者コーパスの2種類が考えられる。また、「書き言葉」と「話し言葉」ではかなり性質が違うので、それぞれをこの2つに分けることができよう。本研究は複合動詞の話し言葉における使用状況に焦点を当てるため、大曾(2006)から抜粋して現在利用できるものの中で、比較的入手しやすい話し言葉コーパスを紹介する。

#### 3.2.1. 日本語母語話者の話し言葉コーパス

「インタビュー形式による日本語会話データベース（上村コーパス<sup>4</sup>）」には、OPI (Oral Proficiency Interview)<sup>5</sup> テスターが日本語母語話者54人に対して行った15分間の日本語 OPI の文字化テキストが収録されており、インターネットで公開されている。このほか、『女性のことは・職場編』、『男性のことは・職場編』に添付されている CD-ROM 所収の会話データがある。また、東京外国語大学 COE の BTS (Basic Transcription System) による多言語話し言葉コーパス<sup>6</sup> の中にも日本語会話データが入っている。以上は会話 / 対話データであるが、独話コーパスとしては国立国語研究所・情報通信研究機構が共同開発した『日本語話し言葉コーパス』<sup>7</sup>がある。

#### 3.2.2. 日本語学習者の話し言葉コーパス

KY コーパス<sup>8</sup>は、OPI に基づいてレベルの付された90人の学習者のインタビューデータで構成されている。また、科学研究費補助金基盤研究(B) (2) 「研究留学生にみられる日本語発話能力の変化と日本語使用環境に関する基礎研究」(研究代表者：尾崎明人) で収集された学習者データの一部、同じく科研費による「就労を目的として滞在する外国人の日本語習得過程と習得にかかわる要因の多角的研究」(研究代表者：土岐哲) で収集されたブラジル人8名の発話資料が名古屋大学留学生センターのサーバーで公開されている。

3.2.1. で紹介した上村コーパスには非母語話者66名のデータが入っている。同じく3.2.1. で触

れた東京外国語大学の話し言葉コーパスの中には母語話者同士の会話に加えて日本人と学習者の会話も収録されている。

### 3.3. 学習者コーパスを利用した日本語習得研究

学習者コーパスを利用した複合動詞習得研究はないため、ここでは、本研究と同じKYコーパスを利用した語彙習得研究の中でKYコーパスを対象に、日本語形態素解析システム「茶釜」<sup>9</sup>（以下「茶釜」と記す）を使った山内氏の研究を取り上げる。

山内(2004)は第二言語としての日本語の語彙習得研究を見渡した論文である。その中で、語彙習得研究においては、研究の絶対量が不足していること、そして、語彙習得研究の中心的な研究である「語の集まり」の習得研究はほとんど扱われていないことを指摘した。次に、語彙習得研究の方法について最近の研究動向を概観した。語彙習得研究の性質上、膨大な量のデータを扱わざるを得なくなる場合があること、現時点ではコンピュータを用いてデータを分析した研究の数は非常に少ないことを指摘している。

山内(2003)は、実際にコーパスを利用した研究であり、「茶釜」を使い、KYコーパスから、日本語能力のレベルを示す特定の形態素を見つけ出す作業を試みた。KYコーパスの中で、連体詞としての「あの」やフィラー「あのー」が出現すればそのデータは「中級」以上であること、助動詞「だ」や終助詞「よ」が出現すれば上級以上であることが決定付けられそうであると述べた。山内(2003)は「語の集まり」の習得を見ていないが、語彙習得研究の一つの方向性を示していると思われる。本研究は山内(2003)の知見を用い、「茶釜」を用いた分析手法を使用した。

## 4. 本研究で使用するコーパス

### 4.1. コーパスの選定理由

学習者と母語話者の話し言葉における複合動詞の使用状況を比較するためには、ほぼ同じ内容、性質の二つのコーパスが望まれる。現在利用できる話し言葉コーパスの中から、コーパスの形式、内容、分量などを総合的に考え、母語話者コーパスとして上村コーパスを、学習者コーパスとしてKYコーパスを用いた。理由は以下の通りである。

#### 1) 形式の同一性

この二つの会話データの収録は、いずれも同じACTFL（全米外国語教育協会）のOPI（口頭表現能力インタビュー）テストの形式に沿って行ったものであり、形式の統一性がとれるため、比較に適している。

#### 2) 分量と話題の類似性

この二つのコーパスは、ほぼ同規模なので、量的分析に適している。また、話題の多様性も考えなければならない。複合動詞の使用は話題のトピックに左右されると考えられるので、二つのコーパスのトピックについて調べたところ、両コーパスとも話題のトピックが多彩であり、トピ

ックによる影響がほとんどないことが確認できたため、これらを用いて比較できると判定した。

### 3) レベル判定の明確性及びデータ構成の良さ

KY コーパスには中, 英, 韓三ヶ国語を母語とする学習者30人ずつ（それぞれ, 初級5人, 中級10人, 上級10人, 超級5人）のデータがバランスよく収集されている。データ構成の整合性が良く, 熟達度別, 母語別の比較など多角的な分析が行える。上村コーパスにも非母語話者66人分の会話データが収録されている。学習者と母語話者の量的比較をしようとするなら, 上村コーパスのみの日本語学習者と母語話者の比較の方が整合性が取れると思われるが, 上村コーパスは学習者の母語と言語背景の構成の記述がなく, 熟達度, 母語別の比較が困難なので, 今回は学習者コーパスを上村コーパスの非母語話者部分のデータではなく, KY コーパスに求めることにした。

以上の理由で, 本研究は母語話者コーパスとして上村コーパスの母語話者50人分のデータ, 学習者コーパスとしてKY コーパスの全データを, 研究利用許諾を得て使用することにした。

## 4.2. コーパスの概要

本研究で利用する二つのコーパスの概要を表1にまとめ, ファイル数, 総文字数などの内訳を表2に示す。

表1 本稿で取り上げる両コーパスの概要

	形式	データの内容	本稿での使用	利点
上村コーパス	OPI形式による日本語会話データベース	日本語母語話者(54人)15分間の簡易OPI	母語話者50人分	①被験者の多様性(年齢層10代~60代) ②形式の統一性(導入部→会話とロールプレイ→終結部) ③発話単位の認定が容易, 数量化, 定量化を行いやすい。
		英語, 韓国語, 中国語, ロシア語などを母語とする66名のOPI	使用しない	
KYコーパス	90人分のOPIテープを文字化した言語資料	中国語, 韓国語, 英語を母語とする30名ずつ, 計90名の学習者の30分程度のOPI	学習者90人分(学習者の発話のみ使用)	①能力レベルが明示されている。 ②インタビューの構成がしっかりしている。データ同士の比較が容易である。 ③発話単位の認定が容易, 数量化, 定量化を行いやすい。 ④中, 英, 韓三国語30人(それぞれ, 初級5人, 中級10人, 上級10人, 超級5人)のデータがあり, 整合性がある。

表2 使用した両コーパスの内訳

コーパス	対象データ	総文字数				
上村コーパス	日本語母語話者 50人分	318,276				
KYコーパス	レベル別分類 ファイル数	初級 5人分	中級 10人分	上級 10人分	超級 5人分	小計
	中国語母語話者 30人分	5,880	43,414	56,962	27,207	133,463
	韓国語母語話者 30人分	6,673	35,441	49,854	35,016	126,984
	英語母語話者 30人分	6,315	36,131	57,873	30,141	130,460
		390,907				

## 5. データ処理手順

複合動詞及びその例文を抽出する前に、まず、上村コーパス、KY コーパスにあるファイルを以下の手順で処理した。作業は正規表現によるテキストエディタでの編集とプログラミング言語 Perl の併用で行った。

### 5.1. データの整形

- 1) 上村コーパスについて、ネットで公開された50人分の母語話者データを統合した。データに混在している OPI テスターのあいづち的な発話、非言語的な記録、文頭の話者コードなどを削除した。
- 2) KY コーパスに対しても、90ファイル中の学習者の発話をそれぞれ抽出した。学習者発話データに混在している OPI テスターの相づち的な発話、非言語的な記録、記録不可能である箇所の表記記号、及び文頭の話者コードなどを削除した。熟達度別に初級、中級、上級、超級とグループ分けし、それぞれ保存した。母語別に、中国語、韓国語、英語に分けて保存した。全てのファイルを統合したデータも作った。
- 3) 両コーパスの総文字数<sup>10</sup>をそれぞれ算出した。KY コーパスのグループごとの総文字数も算出した。話者による繰り返し、言いよどみ、フィラーなどを総文字数に含めている。

### 5.2. 複合動詞の抽出

- 1) 「茶釜」で両コーパスに対し品詞情報を付与した。
- 2) タグ付けされたデータについて複合動詞を抽出<sup>11</sup>し、元のデータと照らし合わせ、複合動詞の一覧表を作成した。そして、全複合動詞を前項動詞別、後項動詞別に集計を行った。KY コーパスの場合、全複合動詞を前項動詞別、後項動詞別、熟達度別、母語別に集計を行った。

## 6. 調査結果と考察

### 6.1. 複合動詞全体使用の比較

#### 6.1.1. 動詞、複合動詞全体使用の比較

全体の使用状況を見るために、二つのコーパスの総文字数、動詞使用数、複合動詞使用数、異なり複合動詞使用数を表3にまとめた。ただし、これらの結果は、どちらもコーパスの規模が30万~40万字程度で、英文コーパスと比較すると必ずしも充分大きいとは言い難いので、結果についても慎重に扱うべきであることは言うまでもないが、種々の傾向を知るには有用であると思われる。

学習者と母語話者における動詞使用数について $\chi^2$ 検定の結果、両者に有意な差はなかった( $\chi^2=2.20, ns$ )。つまり、全体としての動詞使用数に関して、学習者は母語話者とほぼ同程度使用していると言える。しかし、複合動詞使用数対総文字数の比率について、学習者の複合動詞の



使用率 (0.0512%) は母語話者使用率 (0.108%) の約半分しかない。実際、 $\chi^2$ 検定の結果、学習者による複合動詞使用数と異なり複合動詞使用数は統計的に有意に少なかった (複合動詞使用数:  $\chi^2=64.87$ ,  $p<.01$ , 異なり複合動詞使用数:  $\chi^2=30.583$ ,  $p<.01$ )。この結果から学習者による複合動詞使用は頻度も種類も母語話者より有意に少なく、複合動詞は学習者にとって一つの習得困難な語彙項目であることが分かる。この結果は先行研究の指摘の裏づけになると同時に、複合動詞習得研究の必要性を示している。

表3 両コーパスにおける動詞、複合動詞使用の概要

	母語話者コーパス(上村)	学習者コーパス(KY)
総文字数	318,276	390,907
動詞使用数 <sup>12</sup> (総文字数に対する比率)	17,380(5.46%)	21,891(5.06%)
複合動詞使用数 (総文字数に対する比率)	343(0.108%)	200(0.0512%)
異なり複合動詞使用数	174	105

#### 6.1.2. 複合動詞使用上位15項目の比較

母語話者コーパスと学習者コーパスにおける頻出複合動詞の上位15項目とそれぞれの使用数の合計、総複合動詞使用数に対する使用率を表4に示す。

表4を見てみると、母語話者の上位15項目の複合動詞ののべ使用数合計は122語で、母語話者総複合動詞使用数343の35.5%を占めるのに対し、学習者の上位15項目ののべ使用数合計は81語で、学習者総複合動詞使用数200の40.5%を占めている。上位15項目は、母語話者の総複合動詞使用数の343語、学習者の200語のそれぞれ30~40%程度を占めていることから、頻繁に使われる複合動詞グループがあることが分かる。

上位15項目のうち、「付き合う」、「出かける」、「落ち着く」の3項目が母語話者と学習者で共通している(表4の中の下線の項目)。母語話者は「出かける」を25回使用しているのに対し、学習者は8回使用し、「出かける」に関しては、母語話者より使用が少ない。母語話者は「付き合う」を8回使用しているのに対し、学習者は16回使用している。つまり「付き合う」に関して、母語話者より多く使用している。母語話者は「落ち着く」を4回、学習者は5回使用し、ほぼ同程度で使用しているなど、学習者と母語話者の間で使用のばらつきが存在する。

共通している後項動詞を詳しく見てみると、学習者上位15項目のうち、後項動詞「~あう」は4項目合計27回であり、母語話者の2項目14回より有意に多く使われている( $\chi^2=13.51$ ,  $p<.01$ )。すなわち、「~あう」に関して学習者は多く使用しているといえる。その一方、母語話者上位15の項目のうち、「~あげる」を2項目12回使用しているのに対し、学習者は「~あげる」を1項目4回使用している。しかも、母語話者が使用した「~あげる」は「申し上げる」と「差し上げる」であり、いずれも敬語表現であるが、学習者は「取り上げる」を使用し、母語話者とは違った「移動」の意味を付加する機能を持っている「~あげる」を使っている。つまり、学習

表4 全体における上位15項目の使用

母語話者コーパス (上村)	複合動詞 使用数 (総数343)	総複合動詞使用 数に対する使用 率 (%)	学習者コーパス (KY)	複合動詞 使用数 (総数200)	総複合動詞使用 数に対する使用 率 (%)
出かける	25	7.29	付き合う	16	8.00
引っ越す	21	6.12	出かける	8	4.00
働きすぎる	9	2.62	乗り換える	8	4.00
付き合う	8	2.33	受け入れる	6	3.00
見つける	7	2.04	落ち着く	5	2.50
申し上げる	7	2.04	思い出す	5	2.50
打ち込む	6	1.75	取り組む	5	2.50
～し合う	6	1.75	落ち込む	4	2.00
恐れ入る	5	1.46	取り上げる	4	2.00
差し上げる	5	1.46	似合う	4	2.00
話し掛ける	5	1.46	話し合う	4	2.00
結びつく	5	1.46	思い切る	3	1.50
申し込む	5	1.46	聞き取る	3	1.50
落ち着く	4	1.17	繰り返す	3	1.50
思い始める	4	1.17	知り合う	3	1.50
計	122	35.5%	計	81	40.5%

者は「～あげる」に関して使用が少なく、しかも、敬語表現としての複合動詞使用は少ない傾向があるといえよう。したがって、同じ後項動詞である「～あげる」でも、結合する前項動詞によって、意味が多様であり、複合動詞の個別項目の使用頻度を比べるだけでは学習者の使用状況は必ずしも正しく把握できない。よって、以下で、複合動詞構成要素の後項動詞別、前項動詞別でより詳しく見ていく。

## 6.2. 後項動詞別の比較

学習者の使用状況をより詳しく分析するため、母語話者コーパスと学習者コーパスにおける複合動詞全体を、前項動詞別、後項動詞別に分けて比較する。先に、複合動詞の意味づけの中心的役割を果たすことの多い後項動詞を分析する。表5に、後項動詞の使用頻度上位15項目それぞれの使用数、及び種類（ここでは結合する前項動詞の数のこと）、総複合動詞使用数に対する比率をまとめた。

学習者と母語話者の使用した後項動詞上位15項目のうち、10項目は共通している（表5の中の下線の項目）。使用頻度に差はあるが、学習者と母語話者とで使用する後項動詞の大部分は共通している。

また、後項動詞上位15項目の使用数の合計、及び総複合動詞使用数に対する割合は、学習者が

表5 後項動詞の使用頻度上位15位

母語話者コーパス (上村)			学習者コーパス (KY)		
後項動詞 (複合動詞種類) 全種類68	後項動詞 使用数 (総数343)	総複合動詞使用 数に対する使用 率(%)	後項動詞 (複合動詞種類) 全種類43	後項動詞 使用数 (総数200)	総複合動詞使用 数に対する使用 率(%)
かける (6)	34	15.18	あう (12)	38	19.00
あう (11)	30	13.39	だす (13)	20	10.00
こむ (12)	23	10.27	こむ (8)	16	8.00
こす (1)	21	9.38	かえる (5)	14	7.00
あげる (7)	19	8.48	あげる (5)	10	5.00
すぎる (7)	16	7.14	かける (3)	10	5.00
つく (4)	13	5.80	入れる (2)	8	4.00
つける (7)	13	5.80	すぎる (7)	7	3.50
だす (8)	12	5.36	つく (3)	7	3.50
はじめる (7)	12	5.36	つける (4)	6	3.00
きる (7)	11	4.91	くむ (1)	5	2.50
なおす (7)	10	4.46	こす (3)	5	2.50
いる (5)	8	3.57	あがる (3)	4	2.00
あがる (5)	6	2.68	かえす (1)	3	1.50
つづける (6)	6	2.68	かわる (2)	3	1.50
計	234	68.22%	計	156	78%

156回、78.00%であり、母語話者は234回、68.22%である。つまり、母語話者、学習者ともに、後項動詞上位15項目の合計が複合動詞全体使用の7割程度を占めており、この点も両者は類似の傾向を示していると言える。

結合する前項動詞の異なり数を数えた、後項動詞の種類合計を比べると、母語話者68に対し、学習者は43と少ない。これは、先の表3で複合動詞全体における異なりの複合動詞の使用数の差に現れていた違いと同じである。しかしながら、ここで個々に後項動詞を比較すると、種類に差のあるものとなないものがあることが分かる。

たとえば、学習者の上位1位、2位にある項目「～あう」、「～だす」は、それぞれ12種類38回、13種類20回で、母語話者の当該項目使用数はそれぞれ11種類30回、8種類12回である。ほぼ似ている、あるいはやや学習者の方が種類と回数が多い傾向を示している、と言えるだろう。他方、母語話者の1位「～かける」は6種類34回であるのに対し、学習者は3種類10回と、種類も使用数も母語話者より少ない。また、母語話者の4位「～こす」は1種類21回であるのに対し、学習者は3種類5回と、傾向がかなり異なる。母語話者は「引越す」の1種類のみで21回であるのだが、学習者は「引越す」、「追い越す」、「通り越す」の3種類で、合計5回の使用である。このように、個々にみていくと、学習者と母語話者とで複合動詞の使用傾向には違いのある

ことが分かる。

次に、後項動詞のタイプ別に比較する。母語話者、学習者の共通4項目「～あげる」、「～あがる」、「～こむ」、「～だす」と、母語話者の「～いる」、学習者の「～入れる」は、基本義が方向移動を表す後項動詞である。母語話者も学習者も、この方向移動を表す後項動詞はよく用いているということが分かる。

ところが、アスペクトを表す後項動詞に注目すると、学習者は母語話者よりその使用が少ない。母語話者における「～はじめる」、「～つづける」などが、それぞれ7種類12回、6種類6回の使用であるのに対し、学習者による「～はじめる」と「～つづける」の使用は上位15項目に入っていない。アスペクトを表す後項動詞の総複合動詞使用数に対する比率は、母語話者の12.10%に対し、学習者は7.00%である。なお、「～かける」、「～だす」、「～きる」、「～あげる」などは「始動」、「完了」のアスペクトを表す用法以外の事象を表すこともあるため、アスペクトとしての使用例のみ比率に直して計算した。

### 6.3. 前項動詞別の比較

前項動詞別の使用状況を表6に示す。後項動詞と同じく、前項動詞の上位15項目のそれぞれの使用数、及び種類（ここでは結合する後項動詞の数のこと）、総複合動詞使用数に対する比率をまとめている。

学習者と母語話者の使用した前項動詞上位15項目のうち、10項目は共通している。学習者は使用頻度の差はあるが、母語話者と大部分が共通している前項動詞を使用していると言える。

前項動詞上位15項目の使用数の合計、及び総複合動詞使用数に対する割合を比べてみると、母語話者は217回、63.27%であり、学習者は122回、61.00%であるので、前項動詞上位15項目は両コーパスともに、複合動詞全体使用の6割を占めている。

表6 前項動詞別の使用頻度上位15位

母語話者コーパス (上村)			学習者コーパス (KY)		
前項動詞 (複合動詞種類) 全種類80	前項動詞 使用数 (総数343)	総複合動詞使用 数に対する使用 率 (%)	前項動詞 (複合動詞種類) 全種類62	前項動詞 使用数 (総数200)	総複合動詞使用 数に対する使用 率 (%)
<u>で</u> (3)	29	12.95	<u>つき</u> (3)	19	9.50
<u>ひき</u> (7)	27	12.05	<u>とり</u> (8)	16	8.00
<u>とり</u> (14)	24	10.71	<u>で</u> (3)	10	5.00
<u>み</u> (11)	21	9.38	<u>おち</u> (2)	9	4.50
<u>おもい</u> (9)	17	7.59	<u>おもい</u> (3)	9	4.50
<u>し</u> (10)	16	7.14	<u>のり</u> (1)	8	4.00
<u>つき</u> (3)	13	5.80	<u>ひき</u> (4)	8	4.00
<u>もうし</u> (2)	12	5.36	<u>み</u> (6)	8	4.00
<u>さし</u> (4)	9	4.02	<u>うけ</u> (2)	7	3.50
<u>うち</u> (3)	9	4.02	<u>ふり</u> (2)	6	3.00
<u>はなし</u> (2)	9	4.02	<u>たち</u> (3)	5	2.50
<u>はたらき</u> (1)	9	4.02	<u>はなし</u> (2)	5	2.50
<u>たち</u> (5)	8	3.57	<u>いい</u> (3)	4	2.00
<u>よみ</u> (5)	8	3.57	<u>かんがえ</u> (3)	4	2.00
<u>うけ</u> (5)	6	2.68	<u>し</u> (3)	4	2.00
計	217	63.27%	計	122	61.00%

結合する後項動詞の異なり数を数えた。前項動詞の種類合計は、全体、後項動詞の傾向と同じく、母語話者80に対し、学習者は62と少ない。が、個々に前項動詞を比較してみても、後項動詞のように目立つ差は見られない。

たとえば、母語話者において種類数の多いものは、順に、「とり～」の14種類24回、「み～」の11種類21回、「し～」の10種類16回、「おもい～」の9種類17回である。これらは、学習者では「とり～」8種類16回、「み～」6種類8回、「し～」3種類4回、「おもい～」3種類9回と、母語話者に比べると数そのものは少ないが、学習者の使用においては上位にほぼ位置づけられるため、使用傾向としては差がないと言えそうである。

#### 6.4. 学習者熟達度別の比較

学習者の熟達度別の使用状況を比較してみる。学習者の熟達度別及び母語話者の複合動詞使用状況は表7にまとめる。

表7 学習者の熟達度別の複合動詞使用と母語話者の比較

	初級 15人	中級 30人	上級 30人	超級 15人	合計 90人	母語話者 50人
総文字数 (全体に占める比率)	18,868 (4.83%)	114,986 (29.42%)	164,689 (42.13%)	92,364 (23.63%)	390,907	318,276
複合動詞使用数 (総複合動詞数に占める比率)	1 (0.5%)	37 (18.5%)	81 (40.5%)	81 (40.5%)	200	343
異なり複合動詞数	1	20	56	57	105	174

表7の総複合動詞数に占める比率で分かるように、学習者の熟達度別による使用頻度には歴然とした差がある。初級学習者15人による複合動詞の使用は一回のみである。中級学習者30人は20種類37回、学習者の総複合動詞使用数の18.5%を占め、上級学習者30人は56種類81回、40.5%使用している。超級学習者15人による使用は57種類81回、40.5%使用している。上級、超級学習者45人を合わせた複合動詞使用は学習者の総複合動詞使用数の81%にもなる。したがって、学習者による複合動詞の使用頻度は熟達度と相関があり、初級、中級学習者による使用が少なく、上級、超級になるにつれて、使用する複合動詞の頻度も種類も増えていくことが窺える。このことから、複合動詞使用数はある程度、学習者のレベル判定の一つの指標になると考えられる。

学習者の熟達度別による複合動詞使用の特徴をより深く分析するため、学習者熟達度別と母語話者の後項動詞の使用数上位5位をまとめた。初級学習者による複合動詞の使用は「聞き取る」の一語しかなかったため、初級を省いて、中級、上級、超級学習者と母語話者の使用頻度を表8に示す。後項動詞の後の括弧内の数字はその後項動詞で産出した複合動詞の種類の数であり、習得が進むにつれて増えている。なお、「～こす」は1種類21回であったため、省いた。ファイル数が異なるため、直接数字で比べることはしないが、傾向を探るための参考としてみていくことにする。

表8で分かるように、「～あう」、「～こむ」の2項目は学習者の各熟達度のグループと母語話者に共通している。「～あう」の使用頻度は上級、超級学習者のグループの1位になっている。

表8 学習者熟達度別における後項動詞上位5位

中級学習者 (30人)		上級学習者 (30人)		超級学習者 (15人)		母語話者 (50人)	
後項動詞	総数	後項動詞	総数	後項動詞	総数	後項動詞	総数
かえる (2)	6	あう (6)	12	あう (10)	21	かける (6)	34
あう (3)	5	だす (5)	7	だす (7)	9	あう (11)	30
つく (2)	5	こむ (3)	6	こむ (3)	6	こむ (12)	23
だす (4)	4	いれる (2)	6	あげる (3)	4	あげる (6)	19
こむ (2)	4	かえる (3)	5	すぎる (4)	4	すぎる (7)	16

「～あう」, 「～こむ」は学習者の各グループも母語話者も多く使っていることが窺える。超級学習者と母語話者で後項動詞4項目が共通して上位5位に入っている。母語話者がよく用いる「～あげる」, 「～すぎる」は超級学習者も多く使用している。しかも, 「～あげる」の4回のうち, 「申し上げる」が2回を占めている。6.1.2.で学習者は「～あげる」に関して使用が少なく, しかも, 敬語表現としての複合動詞使用は少ない傾向があると記述したが, 超級学習者による「～あげる」の使用は多く, 特に敬語表現の「申し上げる」を使用していることで, 超級学習者は母語話者が用いる敬語表現も使うほどに習得が進んでいると思われる。

## 6.5. 学習者母語別の比較

最後に, 学習者コーパスにおける母語別の複合動詞使用を分析する。各グループの総文字数, 全コーパス総文字数に対する比率, 複合動詞使用数と異なり複合動詞使用数を表9に示す。

表9 母語別の複合動詞使用と母語話者の比較

	中国語母語話者 (30人)		韓国語母語話者 (30人)		英語母語話者 (30人)		合計 (90人)	日本語 母語話者
総文字数	133,463	34.14%	126,984	32.48%	130,460	33.37%	390,907	318,276
複合動詞使用数	78	39.00%	76	38.00%	46	23.00%	200	343
異なり複合動詞数	53	50.48%	46	43.81%	34	32.38%	105	174

中国語, 韓国語, 英語という異なった母語背景を持つ学習者の間で, それぞれ複合動詞使用を比較する。各グループの文字数と全コーパス総文字数の比率はいずれもほぼ33%であり, 有意差がない。しかし, 複合動詞使用には大きな差が見られる。中国語, 韓国語母語話者による使用はそれぞれ53種類78回, 46種類76回と, ほぼ同じ程度であるのに対し, 英語母語話者による複合動詞使用は34種類46語であり, 使用頻度は有意に少なく, 種類も少ないという有意傾向があった。(使用頻度:  $\chi^2(2) = 9.640$ ,  $p < .01$ , 種類:  $\chi^2$ の影響は根強く残り, 不可避的な問題であろう)。その一例として, 対照研究の成果より, 中国語, 韓国語は, いずれも「動詞+動詞」の複合形が許される言語であるのに対し, 英語は「動詞+動詞」の複合形が成立しない言語である。日本語の複合動詞は, 英語の場合には通常前置詞で表すことになる。たとえば, 「持ち出す」は英語では「take out」, 「飛び込む」は「fly/burst in」である。このような言語の相違点が英語母語話者による複合動詞の使用が少ない原因の一つであると考えられる。もちろん, 母語の違いが使用(状況)に影響するかについて, 量的分析のみで簡単には言い切れないが, 今後母語別による質的分析が希求される。

表10に母語別で後項動詞の使用頻度の上位5位を示した。学習者の上位の後項動詞は「～あう」, 「～だす」, 「～こむ」などで母語による違いは無い。また, 母語話者で異なり複合動詞の種類が多い「～あう」, 「～こむ」に対して「～あう」は学習者も種類が多く使用されているが, 「～こむ」は「～あう」ほどではなく, 陳(2004)で指摘している両者の理解の困難度の違いを表

表10 母語別における後項動詞上位5位

中国語母語話者 (30人)		韓国語母語話者 (30人)		英語母語話者 (30人)		母語話者 (50人)	
後項動詞	総数	後項動詞	総数	後項動詞	総数	後項動詞	総数
あう (9)	17	あう (8)	17	こむ (3)	6	かける (6)	34
かえる (3)	7	だす (8)	10	だす (5)	6	あう (11)	30
こむ (2)	4	こむ (4)	6	あう (3)	4	こむ (12)	23
だす (4)	4	いれる (2)	5	かえる (3)	4	あげる (6)	19
あげる (4)	4	かける (3)	4	つける (2)	3	すぎる (7)	16

れていると考えられる。

## 7. 調査結果のまとめ

複合動詞の使用状況を学習者コーパスと母語話者コーパスとで比較しながら分析することにより、学習者の話し言葉での複合動詞の使用状況を明らかにすることができた。今回の調査で明らかになったことを以下にまとめる。

### 1) 全体の比較結果

学習者は全体としては複合動詞使用の頻度も種類も母語話者と比べ、少ない傾向がある。しかし両コーパスの頻出複合動詞上位15項目の延べ使用数合計が、総複合動詞使用数の30~40%程度を占めていることから、頻繁に使われる複合動詞グループがあることが分かる。これらのグループを優先して教育に導入することが必要である。

後項動詞に注目してみると、一部は「~あう」のような、同一複合動詞を反復して母語話者よりも多く使用する項目もあれば、「~すぎる」、「~あげる」のような、母語話者より少なく使用する項目もある。また、「差し上げる」、「申し上げる」、「恐れ入る」など敬語を表す複合動詞も使用が少ない傾向が見える。

このような今回の結果から総合的に判断して、やはり複合動詞は学習者にとって一つの習得困難な語彙項目であることが分かる。この結果は、先行研究の指摘の裏づけになると同時に、複合動詞習得研究の必要性も示している。

### 2) 後項動詞別の比較結果

学習者と母語話者の使用した後項動詞上位15項目のうち、10項目は共通している。学習者は使用頻度の差はあるが、母語話者と過半数は共通している後項動詞を使用しているとは言える。母語話者にとっても学習者にとっても、後項動詞上位15項目は複合動詞全体使用数の7割程度も占めている。複合動詞の教育や習得支援を考える際、これらの項目に重点を置いて行う初期指導が必要である。



学習者は「～あう」、「～だす」のように母語話者よりも多く使用する後項動詞項目もあるが、「～かける」、「～あげる」のような項目は母語話者より少なく使用している。全体として、母語話者より学習者はアスペクトを表す後項動詞の使用が少ない傾向が見られる。

### 3) 前項動詞別の比較結果

学習者と母語話者の使用した複合動詞前項動詞上位15項目のうち、10項目は共通した。この点でも、学習者は使用頻度の差はあるが、母語話者と約2/3は共通の前項動詞を使用していると言える。前項動詞上位15項目は両コーパスともに、総複合動詞使用数の6割を占めている。複合動詞の教育や習得支援を考える際、主に使用する後項動詞だけではなく、使用頻度の高い前項動詞に関する考慮も必要である。

### 4) 熟達度別の比較結果

学習者による複合動詞の使用頻度は習熟度と相関があり、初級、中級学習者による使用は少なく、上級、超級になるにつれて、使用頻度は増えてくるといえる。後項動詞上位5位の比較から、超級学習者は母語話者と比較的近い使用傾向があることが分かった。複合動詞使用数はある程度、学習者のレベル判定の一つの指標になると考えられる。

### 5) 母語別の比較結果

母語別による複合動詞の使用には大きな差が見られた。中国語、韓国語母語話者による総使用数は78語（53種類）と76語（46種類）ではほぼ同じ頻度であるのに対し、英語母語話者による使用は46語（34種類）で、使用頻度も種類も有意に少ないことが分かった。後項動詞上位5位の比較結果から、「～あう」、「～こむ」、「～だす」に共通した使用が見られ、母語による使用頻度のばらつきは見られなかった。

## 8. コーパス利用上の問題点と今後の改善点の提案

言語研究に求められるコーパスとは、単なるデータの蓄積による量の多さだけではなく、質の高さも重要であると考えている。今回、上村コーパスとKYコーパスの使用に際し、いくつかの問題点を感じた。そこで、使用したコーパス、特にKYコーパスへの改善点の提案を通じ、今後のコーパス整備において期待されるべき点について考察する。

### 1) 文字化方針の統一

話し言葉コーパスを構築する際、まず考えなければならないことは文字化の問題である。今回使った二つのコーパスともコーパス内で文字化の表記が統一されていないところがあり、データ整形を行った時作業効率が悪かった。またコーパス内でタグが統一されていない部分があったため、正規表現で一括処理できず、手作業で修正を加えたところが多かった。

今回はコーパスを量的に扱った。量的分析には、特に、コーパス自体の規模がそれほど大きく

ない場合、表記上の問題による偏りが出る可能性が大きい。特に、KY コーパスに学習者による相づちやフィラーなどが多く存在している。このように表記が一定していないため、総文字数の計算や統計に支障をきたした。表記方針を、コーパスを構築する前に慎重に決定し、コーパスを作成する研究者に周知する必要がある。

その他に、文字化する際、漢字で書くか仮名で書くかという表記の違いが非常に大きな影響を及ぼしていることが分かった。「茶釜」で品詞情報を付与して分析を行ったが、仮名表記の多い文の解析では不適切な分析結果になる傾向がある。例えば、「にほん」（日本）を「にる / 動詞 ほん / 動詞」, 「ちゅうごく」（中国）を「ちる / 動詞 うごく / 動詞」と解析された例もあった。品詞情報に基づく複合動詞を抽出する際、このような解析ミスが多かったため、今回はすべて手作業でチェックした。

以上のような問題が解決できるように、今後、利用しやすく整備するためには、コーパス内、コーパス間ともにまず文字化の方針を統一することが望ましい。なお、『日本語話し言葉コーパス』では書き起こしデータの表記が統制されている。

## 2) 属性情報の精緻化

適切な分析が行えるために、コーパスデータの属性情報をより精緻化する必要がある。今回使用したKY コーパスは中国語、英語、韓国語を母語とする学習者30人ずつ合計90人分のデータになっていて、その構成は初級5人、中級10人、上級10人、超級5人の内訳になっている。既存の学習者コーパスの中では、このような熟達度別、母語別で比較できる、横断的かつある程度縦断的にも中間言語を分析できるコーパスはまだまだ足りない。この点で、KY コーパスの貢献は極めて大きい。しかし、KY コーパスにも不足な点がある。まず、より詳しく分析するために、学習者情報の精緻化が希求される。同じ母語、同じレベルの学習者でも、学習環境、日本滞在年数が違えば、当然データに影響してくる。したがって、KY コーパスの改良版に、属性情報の精緻化を期待したい。

## 3) 誤用情報の付加

習得研究を行う際、書き起こされた学習者の発話ごとに誤用のある箇所の明示、誤用の分類、修正のしかたなど誤用情報が付けられれば、分析がより便利になると考えられる。誤用情報が付加されたら、より効率よく、幅広く習得研究に利用できると信じている。今後、KY コーパスと上村コーパスの非母語話者データに誤用情報が付加されることを期待する。

## 4) コーパスの量

今回の調査で、まずコーパスの規模がまだ大きくないことを感じた。データのサイズは調査結果の信頼性に大きな影響を及ぼす。量的研究を行う際、ある程度の量（本研究では総文字数という指標で計った）がないと、調べたい情報が見られないので、誤った、或いは不十分な結論を導く危険性がある。今回利用した上村コーパスもKY コーパスも、文字化表記の統一、属性情報の

精緻化、誤用情報の付加を行ったうえ、量の増加も期待される。

## 5) ツールの開発

コーパスを使用して習得研究を行う際、コーパスの質と量の改善が求められることと同時に、ツールの開発も期待される。今回、「茶釜」を用いて分析を行ったが、OPIのような話し言葉データに「茶釜」を用いることは、根源的な限界があると感じた。学習者コーパス、特に話し言葉コーパスには、学習者による誤用が多く存在するため、その誤用のある部分の形態素解析は、どうしてもおかしなものになってしまう。そもそも、母語話者のコーパスでも、話し言葉に対して、「茶釜」など日本語形態素解析ツールでは正確に品詞情報付与ができないところが多い。例えば、笑い声の「ふ」を「ふる / 動詞」、「へ」を「へる / 動詞」と解析された例も見られた。したがって、今後、話し言葉コーパスに適用する形態素解析ツールの開発が期待される。

## 9. 複合動詞習得研究における今後の課題

今回は話し言葉コーパスを用い、学習者の複合動詞の使用状況を母語話者と比較して量的に扱った。今後は今回の結果を踏まえ、質的分析を行っていくことを予定している。具体的には、KY コーパスから抽出した複合動詞の例文を複数名の母語話者に考察してもらい、誤用のパターンの発見と誤用を生じる原因の追究などを中心に分析する。誤用だけではなく、正用、非用についても深く分析していくことによって、学習者の複合動詞習得の過程について考察したい。

その中でも、学習者による「申し上げる」、「差し上げる」など敬語を表す表現の使用が少ないという今回の分析結果について、その原因の質的分析を行う際に、学習者が敬語用法を知らないのか、敬語を使用する場面でないのかを、コーパスのトピック、及び発話者間の関係などを考慮に入れて詳しく分析する必要があると思われる。また、英語母語話者による複合動詞の使用が中国語、韓国語母語話者より少ないという今回の分析結果について、母語の影響にも関係があるのではないかと考えたが、今後は対照研究の知見を援用し、母語の影響について詳細な分析を行いたい。

また、話し言葉だけではなく、書き言葉に関しても、複合動詞の使用状況を母語話者との比較で量的、質的分析を行う予定である。複合動詞の使用状況について話し言葉、書き言葉両面による調査を行ったうえで、複合動詞の有効な指導法や習得支援などに関する研究を行いたい。

## 注

- 1 「統語的複合動詞」とは補文関係を取る複合動詞で、「話し始める」が「話すことを始める」と言い換えられるように前項が後項の目的語（もしくは主語）になるものである。また「語彙的複合動詞」とは「受け取る、書き込む」のように補文をとらないものである。
- 2 mental lexicon というのは、語彙が習得されているという時、頭の中に語彙の何がどのように内在化されているかという、ひとりひとりが持っている脳内辞書とでもいうべきものである。
- 3 コアは語の意味の全体を見渡すことのできる円錐形の頂点のようなものを表す概念であり、典

型, 非典型を問わず全ての用例の背後にある抽象的な概念である。それを図式で表したのがコア図式である。

- 4 上村コーパスの著作権は北九州市立大学上村隆一氏にあり、『平成8-10年度文部省科学研究費補助金特定領域研究「人文科学とコンピュータ」公募研究(「日本語会話データベースの構築と談話分析」研究代表者:上村隆一)の成果によるものである。上村コーパス(母語話者データ)の会話インタビューの形式は会話モードとロールプレイの二つの部分によって構成され、全行程は15分前後である。インタビューは大学生・教職員・主婦等に対し数名の面接者(OPI テスター有資格者)が行った。会話モードでは、自己紹介、専攻分野、仕事、住居などについての社交的会話や社会的、時事的問題なども取り上げている。ロールプレイは、ゴミの捨て方の説明、映画/旅行の誘い、約束の変更、バイトの面接の4種類の場面設定で実施した。
- 5 OPI とは最長30分という限られた時間内のインタビューで、被験者の口頭能力を最大限発揮させ、妥当で信頼性のある自然な発話を必要最大限採集、録音し、それを ACTFL 外国語能力基準に照らし合わせ、被験者の口頭能力を判定する評価法である(鎌田 2006)。
- 6 コーパスの内容や入手方法は、<http://www.tufts.ac.jp/ts/personal/usamiken/corpora2007.htm>を参照されたい。
- 7 コーパスの内容や入手は、<http://www.kokken.go.jp/katsudo/seika/corpus/>を参照されたい。
- 8 KY コーパスは『第2言語としての日本語の習得に関する総合研究』(平成8年度~平成10年度基盤研究(A)(1)課題番号08308019)通称「AJプロジェクト」(研究代表者:カッケンブッシュ寛子)の研究を遂行するために収集した90人分のOPIのテープを文字化した言語資料である。KY コーパスは、留学生に対し、面接は数名の面接者(OPI テスター有資格者)が行い、自己紹介、専攻分野、出身地とその文化、日本での生活、社会的、時事的問題など社交的会話による応答とロールプレイによって構成される。ロールプレイはレベル別に、隣人への苦情、食事の誘い、警察への通報、友人の結婚式でのスピーチなどの場面設定で実施した。
- 9 「茶筌」は奈良先端科学技術大学院大学情報科学研究科自然言語処理講座(松本裕治研究室)で開発された日本語形態素解析ツールの一種である。単語(形態素)の分割、基本形や品詞、活用情報の提示といった、コーパスに基づいた日本語研究を行ううえで重要な情報を提供してくれる有用なツールである。詳細は<http://chasen.naist.jp/hiki/ChaSen/>を参照されたい。
- 10 日本語の場合、スペースで区切られた単語という単位では記述されておらず、また、学習者の文は誤用を含んでいるので、総単語数ではなく、総文字数という単位でデータの規模を示した。
- 11 具体的な抽出手順は滝沢(2004)を参照されたい。「茶筌」で形態素解析する際、複合動詞を一つの動詞として解析された場合も、「動詞+動詞」として解析された場合もある。二通りに抽出作業を行い、「動詞+動詞」の場合、全データを通して手作業で確認した。以下、複合動詞の出現数はすべてこの二つの合計である。
- 12 「茶筌」で品詞情報を付与された二つのコーパスデータに対し、動詞を抽出した。話し言葉コーパスであるため、「茶筌」によって適切に品詞が付加されず、動詞と分析されたもの(「ちゃう」、「てる」、「しょう」など)を取り除いた数である。

#### 参考文献

大曾美恵子(2006)「日本語コーパスと日本語教育」『日本語教育』130, 3-10, 日本語教育学会

- 鎌田修(2006)「KY コーパスと日本語教育研究」『日本語教育』130, 42-51, 日本語教育学会
- 影山太郎(1993)『文法と語形成』 ひつじ書房
- 影山太郎(1996)『動詞意味論—言語と認知の接点』 くろしお出版
- 影山太郎・由本陽子(1997)『語形成と概念構造』 研究社出版
- 現代日本語研究会(1997)『女性のことば・職場編』 ひつじ書房
- 現代日本語研究会(2002)『男性のことば・職場編』 ひつじ書房
- 滝沢直宏(2004)「日本語電子化テキストからのコロケーションの抽出」『日本語学習辞典編纂に向けた電子化コーパス使用によるコロケーション研究』平成13年～15年度科学研究費補助金基盤研究(B)(2) 報告論文集, 27-40
- 陳曦(2004)「中国人学習者における複合動詞の習得に関する一考察—「～あう」と「～こむ」の理解に基づいて—」『ことばの科学』17, 59-79, 名古屋大学言語文化研究会
- 寺田裕子(2001)「日本語の二類の複合動詞の習得」『日本語教育』109, 20-29, 日本語教育学会
- 姫野昌子(1999)『複合動詞の構造と意味用法』 ひつじ書房
- 松田文子(2002a)「複合動詞研究の概観とその展望—日本語教育の視点からの考察—」『言語文化と日本語教育 5月特集号 第二言語習得・教育の最前線—あすの日本語教育への道しるべ』, 170-184, お茶の水女子大学日本語言語文化学会
- 松田文子(2002b)「日本語学習者による『～こむ』の習得」『世界の日本語教育』12, 43-59, 国際交流基金日本語国際センター
- 松田文子(2004)『日本語複合動詞の習得研究—認知意味論による意味分析を通じて』 ひつじ書房
- 山内博之(2003)「OPI データの形態素解析—判定基準の客観化・簡易化に向けて—」『実践女子大学文学部紀要』45, 1-10, 実践女子大学文学部
- 山内博之(2004)「語彙習得研究の方法—茶釜と N グラム統計」『第二言語としての日本語の習得研究』7, 141-162, 凡人社
- Nesselhauf, N.(2004) "Learner corpora and their potential for language teaching". In Sinclair, John McH. (ed.), *How to Use Corpora in Language Teaching*, 125-152. Amsterdam: J. Benjamins.
- Granger, S. et al. (eds.) (2002) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

#### 付 記

名古屋大学の木下徹先生には本稿の執筆にあたりご指導を賜り、また同大学の滝沢直宏先生、杉浦正利先生にも数々のご助言をいただいた。査読者・編集委員会の方々には貴重なコメントを多数いただき、最後まで修正に付き合っていたいただいた。友人の土肥猛さんには日本語の面でアドバイスをいただいた。記して心より感謝申し上げます。

(投稿受理日：2007年1月31日)

(最終原稿受理日：2007年7月26日)

---

陳 曦 (ちん き)

名古屋大学大学院国際開発研究科国際コミュニケーション専攻  
464-8601 名古屋市千種区不老町  
m050309d@mbox.nagoya-u.ac.jp

# A comparison of Japanese compound verb usage by native speakers and learners of Japanese as a second language: A corpus-based approach

CHEN Xi

Graduate Student, Nagoya University

## Keywords

compound verb, learner corpus, spoken language corpus, Second Language Acquisition

## Abstract

This paper investigates the usage of Japanese compound verbs spoken by learners of Japanese as a second language (JSL). This study focuses on how to use the corpus for Second Language Acquisition research, utilizing two kinds of corpus, “learner corpus” and “native speaker corpus”, and then compares them to analyze usage. The purposes of this study are to clarify the usage of Japanese compound verbs by JSL learners, and to utilize corpora and identify their problems.

The research findings show that the JSL learners’ general usage of compound verbs was more limited than the Japanese native speakers, not only in terms of number but also in variety. In particular, learners used few compound verbs which express respect such as ‘*sashi-ageru*’, ‘*moushi-ageru*’. There were similarities among the two corpora in the use of the secondly-positioned verbs and the firstly-positioned verbs. On the other hand, JSL learners used compound verbs such as ‘*-au*’, ‘*-dasu*’ more than native speakers but used few other compound verbs which show aspect, such as ‘*-hajimeru*’, ‘*-tuzukeru*’. Furthermore, there was a significant difference in the frequency of use and the kind of compound verb depending on learners’ level and learners’ native language.

A number of problems with existing corpora are identified, including the following: descriptions are inconsistent, information on learners’ personal background is insufficient, and information on learners’ errors is not included. Finally, suggestions are made for improvement of corpora.