

web 公開予定文法用例検索システム「日本語文法項目用例文データベース『はごろも』」のレベル付けと学習者コーパスの比較

堀 恵子 (東洋大学文学部)
江田すみれ (日本女子大学文学部)

Comparison of the Levels of Grammatical Items between Learners' Corpus and "HAGOROMO", a Web-based Searching System of Grammatical Items

Keiko Hori (Faculty of Literature, Toyo University)
Sumire Goda (Faculty of Humanities, Japan Women's University)

1. はじめに

筆者らは、日本語教育の教師支援を目的として、文法項目の用例文を、複数のコーパスから抽出し、web 上で公開する「日本語文法項目用例文データベース『はごろも』」(以下、「はごろも」)を作成している。文法項目数は 1,884 項目である。文法項目には主観判定によって 6 段階のレベル付けを行った(堀ほか 2012)。今後はこの段階付けが妥当か検証する必要がある。本発表では、条件表現のバとタラを含む文法項目を取り上げ、学習者コーパスにおける文法項目の使用レベルと「はごろも」の主観判定によるレベルとを比較し、「はごろも」のレベル付けの妥当性を論じる。

2. 「はごろも」の概要

2.1 研究の背景

日本語教育において、長くシラバスの拠り所となってきたのは、旧日本語能力試験の出題基準(以下、旧出題基準)である。1994 年に公開され、2002 年の改訂を経て、受験する学習者向けの授業だけでなく、教科書をはじめとした教材、種々の試験、研究における難易度の目安としても広く用いられてきた。しかし、日本語能力試験は 2010 年に改定され、出題基準は非公開となった。そこで、教育現場、特に海外の日本語教師の間には、何を教えるべきか不安が広がっていると言われている。

その旧出題基準の文法項目は、江田・小西(2008)、堀ほか(2009)、砂川ほか(2011)によると、3, 4 級, 1 級項目の頻度調査の結果、あまり使われない項目があると指摘されている。そのため、新日本語能力試験の出題基準が今後も公開されない場合に旧出題基準を使いつづけることは、学習者のコミュニケーション能力養成のために適切ではないと言えるであろう。

また近年、日本語教育における文法教育を見直そうという気運が高まっている。言語教育の目的は学習者のコミュニケーションを向上させることであるが、日本語学研究成果に依拠しすぎて、日本語教育として必要でないものまで教えてきたのではないかと指摘されている(野田 2005)。しかしながら、ではどの項目が教えられるべきかについては、個々の項目について、どの用法が初級には不要かという提案は見られる(田中 2005 など)ものの、包括的な初級文法シラバスの提案には至っていない。

そこで、英語における van Ek and Trim (1991) の機能項目一覧のような、学習者のコミュニケーション能力を向上させることに役立つ文法項目一覧の作成が望まれる。しかし、現在学習者が多様化していることから、学習者のニーズも多様化しており、日本語教師が向きあう学習者にとって必要な項目も一様ではない。したがって現場に立つ教師自身が、文

法項目をよく理解し、取捨選択する目を持つことが望まれる。しかしながら、特に海外においては日本語の用例に十分に触れる機会が少ない場合もある。そこで、文法項目の用例や頻度情報、難易度の情報を与える項目一覧の作成が望まれる。

以上のことから、筆者らは主に日本語教師支援を目的として、初級から上級までの包括的な文法項目の一覧を作成し、複数のコーパスから用例を抽出し、海外においても利用しやすいように web 上で検索できるシステムを構築することとし、2010 年から活動を始めた。

2.2 文法項目の選定

文法項目は、これまで広く用いられてきた文献の 2 編以上に取り上げられている項目を中心に選定した。用いた文献と選定の理由は下記の通りである。

- A. 「旧出題基準」：初級から上級までの項目を網羅し、これまで多くの教育現場、研究で参照されてきた。
- B. 『日本語文型辞典(以下、文型辞典)』：多くの文型、複合辞などを見出し語として採用し、意味用法を解説している。国内外の日本語教師にも広く参考にされている。
- C. 『現代語の助詞・助動詞(以下、助詞・助動詞)』：助詞、助動詞を包括的に扱っている。
- D. 『日本語表現文型(以下、表現文型)』：複合辞を包括的に扱っている。
- E. 『現代語複合辞用例集(以下、複合辞)』：複合辞を包括的に扱い、複合辞のカテゴリ分けが合理的である。

以上の文献を参考にして、2012 年 3 月に 1,884 項目を選定し、名称を「はごろも」文法表とした。表 1 は、上記文献と、「はごろも」文法表との共通項目の数である。

表 1 参考文献と「はごろも」文法表との共通項目数 (堀ほか 2012 より)

参考文献	旧出題基準	文型辞典	助詞・助動詞	表現文型	複合辞
共通項目数	881	1,468	425	549	350

「はごろも」文法表では、1 つの形式に対し複数の用法、機能も取り上げる。例えば、「ている」は、旧出題基準では、「動作の継続」と「結果の状態」の例文だけが取り上げられていた。「はごろも」文法表では上記の文献を参考に、日本語教育において重要と思われる下記の 6 つを取り上げることにした。

- (1) 動作作用の継続 「私は今本をよんでいます。」
- (2) 結果の状態 「まどがしまっています。」
- (3) 繰り返しの行為 「ここでは、過去に何度も事故が起こっている」
- (4) 経験 「彼は 1 ヶ月前に会社を辞めている」
- (5) 恒常的な状態 「道が曲がっている」
- (6) 完了 「子供が大学に入るところには、父親はもう定年退職しているだろう」

2.3 用例の収集とデータベースの概要

用例は、話し言葉、書き言葉の複数のコーパスから抽出する。対象コーパスは下記の通りである。

(1) 書き言葉

「日英新聞記事対応付けデータ (JENAAD)」 / 「ブログデータ (京都大学・NTT による)」
 京都大学情報学研究所・NTT コミュニケーション科学基礎研究所 共同研究ユニットによる <http://nlp.kuee.kyoto-u.ac.jp/kuntt/> (以下、ブログ) / 「白書」 / 「CASTEL/J CD-ROM V1.5」日本語教育支援システム研究会 (以下、CASTEL/J) / 「日本語教科書」

(2) 話し言葉

「日本語会話データベース」平成 8 - 10 年度文部省科学研究費補助特定領域研究「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 (以下、「上村コーパス」) / 「宇都宮大学 パラ言語情報研

究向け音声対話データベース (UADB) / 「名大会話コーパス」科学研究費基盤研究 (B) (2) 「日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究」(平成 13 年度～15 年度, 研究代表者: 大曾美恵子) / 「BTS による多言語話し言葉日本語会話 1」宇佐美まゆみ監修(2005) 『BTS による多言語話し言葉コーパス・日本語会話 1』東京外国語大学大学院地域文化研究科 21 世紀 COE プロジェクト「言語運用を基盤とする言語情報学拠点」

用例の抽出と, データベースの作成は, 共同研究者とともにを行った。まず, 複数のコーパスから, 形態素解析に基づく下処理を行った上で, 文字列と品詞情報を組み合わせて用例の候補を抽出する。それを目視で精査して, 用例文データベースを作成する。そして web 上の検索システムによって提供するようにする。

用例文の公開に先立ち, 文法表と次節で述べる文法項目のレベルを付与した文解析システムをすでに公開している。テキストを入力すると, 該当する文法項目と語彙項目について, その候補とレベルを表示する。結果は文法項目と語彙項目に分けて出力され, 集計もできる。また, 解析結果は CSV と HTML の 2 通りの方法で保存できる。

2.4 文法項目のレベル付け

文法項目のリストがあっても, 難易度が分からなければ, どのレベルの学習者に教えるべきか分からない。そこで文法項目に対してレベル付けを行った(堀ほか 2012)。

付与するレベルは, 日本語能力試験のような特定の試験に合わせるのではなく, 初級, 中級, 上級という一般的に行われているレベル感をもとにし, さらにそれを細かくした「初級」¹「初中級」「中級」「中上級」「上級」「超級」の 6 段階とした。CEFR が 6 段階であることも参考にしたが, CEFR の能力記述文に合わせてレベルを位置づけたのではなく, 評定者である日本語教育経験者の判断によって 6 段階に分ける主観判定とした。

評定は, 評定者に文法項目, 文法カテゴリー, 用法・機能, 典型的な例文を記した一覧表を示し, 他の評定者の結果が分からない状態で判定させた。評定者は日本語教育経験が 10 年以上の日本語教師 7 名である。調査期間は 2011 年 12 月から 2012 年 2 月 1 日までであった。

評定の結果, 7 名のうち 5 名の評定について, 平均値との間で中程度の一致が見られたため, 5 名の評定結果を採用し, その平均をもってレベルを決定した。

評定の結果得られた各レベルの文法項目数は, 表 2 の通りである。

表 2 6 段階レベルごとの項目数

「初級」	「初中級」	「中級」	「中上級」	「上級」	「超級」
155	196	351	592	523	67

3. レベル付けの検証

今後は主観判定によって得られたレベルが妥当なものであるかどうかを検証していかなければならない。そのためには, 母語話者の産出データについて, 頻度情報, 使用されるジャンル, レジスターの調査を行い, それとともに, 外的基準となるテスト成績などのレベル付けがある学習者の産出データとの比較などの方法が考えられる。それぞれの項目についてこれらの調査を行うのは膨大な時間がかかることが予想される。

本発表では, その手始めとして, 条件形式バとタラを含む項目について, 口頭能力試験である OPI²の判定が付与されている学習者データを調査し, OPI 判定と「はごろも」文法表のレベル分けを比較する。

¹ 「はごろも」文法表に付与したレベル分けを, 一般に行われている初級, 中級, 上級概念と区別するため, 本稿では「」に入れて示す。

² OPI(Oral Proficiency Interview)は, 最長 30 分でテスターが与えた質問やタスクに受験者が応答し, それを録音してレベルを判定するという形で口頭能力を測るものである。(鎌田 2006)

KY コーパスのテキストから、表 3 に示したバ、タラを含む表現を抽出した。

3.3 結果

その結果、591 例が抽出された (表 4)。自ら言い直しをしている場合は、あとの発言を採用した。下の例 1 は、下線部の後言い直しをしているため、使用例としては採用しない。

例 1) それでちょっと、すいませんが、〈はい〉予約をキャンセルしていただければ、いただきたいんですが。 (KA01)

誤用は、条件形式を使用しているも、意味、機能が不適切な場合は、誤用とした。例 2 は、話題を提示する複合辞として使用しているが、後ろには話題に取り上げたものを説明する完全な形の文が来なければならない。しかし、この例は「花と言えば吉野」のような慣用表現に見られるように完全な形ではなく、すわりの悪い文となっている。そこで、誤用とした。

例 2) Tさんの年齢とぴったりと思いますね、まあ、性格と言えば、うん、優しい男ですよ。 (CA02)

表 4 KY コーパスに見られる条件形式バとタラを含む使用数

	「初中級」 (たら仮定、ば+未実現のことから、ば条件)	「中級」 (そういえば、たら～た、たら～だろう、ば+意志・希望、ば～た/～ていた、ばいい、ばくり返し・習慣、もしかしたら～か)	「中上級」 (からいえば、からみれば、さえ～ば、てみれば、といえ、なければ～ない、なぜかといえ～からだ、ば+働きかけ、ば～ほど、ば前置き、ば立場・観点、疑問詞+～たら～のか、ば～のに)	「上級」 (～も～ば～も～、といえ、といえ～が)	総計
IL	5	1	0	0	6
IM	27	12	1	0	40
IH	40	6	4	0	50
A	51	17	6	1	75
AH	116	58	36	4	214
S	141	40	21	4	206
総計	380	134	68	9	591

3.4 考察

表 4 を見ると、使用数は、総計においても、各レベルの使用数においてもおおむね OPI レベルが上がるにしたがって数が増えている。またその出現は、「初中級」「中級」の項目から、OPI レベルが上がるにしたがって「中上級」「上級」の項目へと広がりが見られ、使用数も増えている。

このことから、6 段階レベル付けは、少なくとも本稿の調査の範囲内では妥当であると言えよう。

今後は検証する項目を増やし、母語話者の産出データとの比較も行っていく。

4. まとめと今後の課題

本稿では、「はごろも」について紹介するとともに、文法項目に付与した 6 段階レベルの妥当性を検証する試みとして、条件形式バとタラを含む項目を取り上げて考察した。学習者のレベルが確定している KY コーパスにおける使用数は、「はごろも」文法表の「初中級」と「中級」レベルの項目から、「中上級」と「上級」の項目へと広がりが見られ、レベルが上がるにしたがって使用数が増えていた。その結果、本稿で取り上げた項目のうち使用が見られた項目については、レベル分けが妥当であると言えよう。

今後の課題として 3 点上げる。第 1 に、本稿で扱った項目は話し言葉だけでなく、書き言葉において使用される項目も含んでいる。したがって、今回使用が見られなかった項目

についても、直ちに使用が少ないと結論づけることはできない。今後は、話し言葉についてはより大きなコーパスを対象として調査する必要がある。また、書き言葉の調査も必要である。第2に、話し言葉のコーパスに関して、堀(2012)は、KYコーパスは口頭能力試験であることから、試験官が能力の上限を見極める目的で言語能力を超えた「突き上げ」と呼ばれる質問をするため、被験者が緊張を強いられることと、被験者が誤りを恐れて難しい項目を使用しない非用が見られる可能性とを指摘している。したがって、学習者の産出データとしては、口頭能力試験以外のデータがより望ましい。第3に、各文法項目について、母語話者データを調査して、母語話者の使用頻度と比較する必要がある。

今後は調査項目と調査対象を広げ、母語話者コーパスとの比較も行い、統計的に6段階のレベルの妥当性を検証していくことが課題である。

謝 辞

本研究は、文部科学省科学研究費補助金基盤研究(C)「日本語教育のためのコーパスに基づく文法項目データベース構築と検索システムの公開」課題番号:22520538、(代表:堀恵子)の助成を受けている。

文 献

- 鎌田修(2006)「KYコーパスと日本語教育研究」『日本語教育』130, pp.42-51.日本語教育学会
- グループジャマシイ(1998)『教師と学習者のための日本語文型辞典』くろしお出版。
- 江田すみれ、小西円(2008)「3種類のコーパスを用いた3級4級文法項目の使用頻度調査」『日本女子大学 紀要 文学部』57, pp.1-28.
- 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。
- 国立国語研究所(1951)『現代語の助詞・助動詞-用法と実例-』秀英出版。
- 国立国語研究所(2001)『現代語複合辞用例集』国立国語研究所。
- 砂川有里子、清水由貴子、奥川育子、千葉庄寿(2011)BCCWJによる機能語データベース(スタンドアロン版)(Ver. 0.9.1b)特定領域研究「日本語コーパス」研究成果報告書
- 田中真理(2005)「学習者の習得を考慮した日本語教育文法」野田尚史編(2005)『コミュニケーションのための日本語教育文法』pp.63-82.くろしお出版。
- 野田尚史編(2005)『コミュニケーションのための日本語教育文法』くろしお出版。
- 堀恵子(2005)「日本語条件表現の習得過程-中級学習者に対する縦断的インタビューから-」『日本語教育方法研究会誌』12:1, pp.36-37.日本語教育方法研究会
- 堀恵子(2007)「日本語条件文の文末制約習得に及ぼす母語の影響-タイ語・英語・韓国語・中国語話者を対象とした文法性判断テストから-」『麗澤大学紀要』84, pp.101-126.麗澤大学
- 堀恵子(2012)「習得過程研究における学習者コーパスの制約-OPIコーパスとインタビューの比較から-」『人間科学総合研究所紀要』14, pp.95-118.東洋大学人間科学総合研究所
- 堀恵子、荒川みどり、小池恵己子、小林佳代子(2009)「日本語能力試験出題基準の<機能語>を対象としたコーパス調査-目標言語使用領域での課題遂行に必要な項目を検証する-」『2009年度日本語教育学会春季大会予稿集』pp.194-199.
- 堀恵子、李在鎬、砂川有里子、今井新悟、江田すみれ(2012)「文法項目の主観判定による6段階レベルづけとその応用」2012年日本語教育国際研究大会ポスター発表
- 森田良行、松木正恵(1989)『日本語表現文型』アルク。
- van Ek, J.A. & Trim, J. L. M. (1991) *Threshold 1990*. Cambridge: Cambridge University Press.

関連 URL

学習項目解析システム <http://lias.intersc.tsukuba.ac.jp/>