

研究資料

機能語用例文データベース「はごろも」について

堀 恵子 (東洋大学/筑波大学)

李 在鎬 (早稲田大学)

長谷部 陽一郎 (同志社大学)

要旨

筆者らは、日本語教師、特に海外の非母語話者日本語教師の支援と、中上級以上の学習者の支援を目指して、日本語機能語用例文データベース「はごろも」を開発した。検索システムでは、日本語教育に必須な機能語と、話し言葉と書き言葉のコーパスから抽出した真正な用例を検索することができる。現在、web 上の検索エンジンと、Excel 表形式のダウンロードの2つの方法で利用できる。見出し語は、旧 JLPT の『出題基準』など5つの資料から選んだ1,848語である。用例は、名大コーパスなど4種類の話し言葉コーパスと、CASTEL/J を含む4種類の書き言葉コーパスから抽出した。項目の難易度は、10年以上の日本語教育経験のある5名の日本語教師による主観判定によって決定された。

キーワード：機能語、web 検索エンジン、真正な用例、話し言葉コーパス、書き言葉コーパス

1. 機能語用例文データベース「はごろも」の背景と目的

1.1 機能語用例文データベース開発の背景と目的

国際交流基金の2012年の調査によると、全世界中の何らかの機関で学んでいる日本語学習者は約399万人であり、海外の日本語教師は64,000人近くに上る(国際交流基金2012)。一方、日本語教育上の問題点として、「教師不足」「教師の日本語能力」「教材不足」「施設・設備不足」「教材・教授法情報不足」が、その前の2009年調査と比較すると改善されてきているとは言え、まだ大きな問題として挙げられている。

日本語教師が文法を教えようとするとき、まず教師自身が当該項目をよく理解し、学習者に対して豊富な用例文の形で形式、意味、機能、使用される表現形態などを示すことが必要である。しかし上記のような問題が指摘されるなか、教師の日本語能力について、荒川・木山(2005)は、国内の機関で研修を受けた非母語話者教師の日本語運用能力は初級後半から上級後半まで幅広く、母語話者教師向け文法参考書を手軽に読んですぐに理解できるとはいいがたいと指摘している。

そこで、日本語教師自身が真正な用例に豊富に触れることができれば、機能語の意味や共起する語、使用されるジャンルなどを理解できると考え、筆者らはこれらの情報をウエ

ブ上で提供することを目指した。それによって、特に海外の非母語話者日本語教師を支援することができるであろう。さらに、中級以上のレベルの学習者の自律学習支援にも役立つと考える（堀・江田・李 2010, 堀・江田 2011, 堀 2012, 堀・江田 2013, 堀・李・江田 2015a, 堀・李・江田 2015b）。

1.2 難易度表記の必要性について

本データベースでは、機能語に難易度を示している。その背景は、以下の2点である。1つめは、長らく日本語教育において教えるべき文法の目安とされてきた日本語能力試験（以下、JLPT）の『日本語能力試験出題基準』（以下、旧『出題基準』）が改定され、非公開となったことである。JLPTは1984年から実施されており、2014年には世界で60万人近くが受験している。1994年に旧『出題基準』が公開されて以来、そこに挙げられた項目はJLPT受験のためだけでなく、より広くシラバスの基準として扱われてきた。また、日本語教育学研究においても参考にされてきた。しかし、2010年のJLPT改定にもなつて新たな出題基準は非公開となり、受験者とともに受験者を指導している教師にも、どのような機能語を教えるべきか不安が広がった。いわゆるJLPTのための対策本では、かなり難解な項目を取り上げるものもある。

2つめは、コミュニケーションのための文法教育が見直されていることである。野田（2005）は、日本語教育においては日本語学に基づいた文法が教えられてきたが、真にコミュニケーションのための文法を教えるべきであると指摘し、これまでの教育の見直しを訴えた。それ以後、新たに何を教えるべきかに関して模索が続いている。

以上の2点から、本データベースでは、機能語のリストを示すだけでなく、どの段階の学習者に何を教えるかの目安となるよう項目の難易度を示すことが必要であると考え、日本語教育経験者による難易度を提示することとした。

1.3 データベースの名称について

本データベースは2010年度より科学研究費の助成を受けて行っており、当初「文法項目用例データベース『はごろも』」と称していた。しかし、2015年3月に用例以外のリストを公開する際に、「機能語用例文データベース『はごろも』」と名称を改めた。機能語とは、文型、表現を含めた文法項目を指すものである。

2. 利用した資料

2.1 機能語の選択

機能語を選択するにあたって、次に示す5種類の資料に収録されている項目の中で、主に2種類以上の資料に収録されている項目を選んだ。下記は5種類の資料である。

- (1) 旧『出題基準』（1994）
- (2) グループジャマシイ（1998）『日本語文型辞典』（以下、『文型辞典』）
- (3) 国立国語研究所（1951）『現代語の助詞・助動詞』（以下、『助詞・助動詞』）
- (4) 森田・松木（1989）『日本語表現文型』（以下、『表現文型』）
- (5) 国立国語研究所（2001）『現代語複合辞用例集』（以下、『複合辞』）

2.2 用例の抽出

用例を抽出したコーパスは、下記に示すように、話し言葉のコーパス4種類と書き言葉4種類である。2つの表現形態から用例を抽出することによって、表現形態の違いによる

使用の偏りを利用者が理解しやすくなることを意図した。

書き言葉

- (1) 「日英新聞記事対応付けデータ (JENAAD) 」
- (2) 「KNB コーパス」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/>
- (3) 「白書」
- (4) 「CASTEL/J CD-ROM V1.5」日本語教育支援システム研究会
 - ▶ bookdata: 主に新書・白書・日本語能力試験の問題と選択枝 (用例としては日本語能力試験の問題と選択枝は除いている)
 - ▶ taradata: 映画『男はつらいよ』のシナリオ

話し言葉

- (1) 「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
- (2) 「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 (UUDB) 」
- (3) 「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)
- (4) 「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」

3. 構築したデータベース

3.1 統計情報

採用した機能語は 1,848 項目である。元にした資料別の機能語数は表 1 のとおりである。各資料に収録されている項目は互いに重複しているため、合計数ははごろもの機能語数より多くなっている。

表 1: 元の資料から取り入れた「はごろも」の項目数

旧『出題基準』	『文型辞典』	『助詞・助動詞』	『表現文型』	『複合辞』
956	1,479	413	597	346

旧『出題基準』はこれまでの影響の大きさから、すべての項目を取り入れている。『文型辞典』からは、初級から上級までの文法項目の多くを取り入れている。『助詞・助動詞』からは助詞と助動詞を、『表現文型』と『複合辞』からは複合辞を取り入れている。

以上のように広く文法に関する項目を複数の資料から取り入れているため、機能語として種類の異なるものが混在している。活用形、慣用的表現、その他の項目である。活用形とは、言語形式の活用形にのみ焦点を当てた項目で、「V ます 動詞の現在形 polite 肯定」「A く 形容詞の副詞用法」などがある。また、慣用的表現には、「ていただけますか」「てもよろしい (ですか)」のような文末表現がある。最後に、その他の項目は、上記

2つ以外の項目である。

3.2 難易度別の項目の分布

項目難易度は、初級前半、初級後半、中級前半、中級後半、上級前半、上級後半の6段階とし、5名の日本語教師の主観判定によって行った(堀ほか2012)。

表2は、「はごろも」収録の機能語のうち、旧『出題基準』にあった項目と、新たに取り入れた項目の分布を示している。表中、旧3級の175項目に注目すると、「はごろも」6段階では初級前半から上級前半までの5段階に振り分けられており、旧3級には幅広い難易度の項目があったと判断されたことが分かる。

旧『出題基準』にあった956項目の級と、「はごろも」6段階とはともに順序性があるため、グッドマン・クラスカルの順序連関係数 γ (ガンマ)を用いたところ、 $\gamma=0.893$ と強い相関が見られた。全体的に見ると、「はごろも」6段階の難易度は旧『出題基準』の級に沿ったものと言える。

表2:「はごろも」の6段階レベル別見出し語数

「はごろも」6段階	旧『出題基準』				旧出題基準にない項目	総計
	4級	3級	2級	1級		
初級前半	140	5			10	155
初級後半	65	66			63	194
中級前半	4	98	70		173	345
中級後半		5	254	18	305	582
上級前半		1	85	118	301	505
上級後半			3	24	40	67
合計	209	175	412	160	892	1,848

次に、旧『出題基準』になく、「はごろも」にある項目の例を示す(表3)。接続詞、副詞は旧『出題基準』の「文法」には含まれていなかったが、「はごろも」には取り入れている。

表3: 旧『出題基準』になく「はごろも」にある見出し語

レベル	項目例	品詞など
初級前半	それから もう少し/もっと ほうが~より (も)	接続詞 副詞 文型
初級後半	しかし/また/そして (のこ)を~という/もしも~たら いい/がほしいんですが	接続詞 文型 表現
中級前半	そういえば/そこで/ところが てもよしい(でしょうか) Vていた/とちがって/ないと/なくて	接続詞 表現 その他
中級後半	および/かつ/くわえて だから~のだ/どうも~そうだ となると/Vなおす/にくわえて	接続詞 文型 その他
上級前半	それはそうと/ところで/ちなみに あえて/あまり(に)/とんだ ~であろうと, ~であろうと でもって/にしても/にかざる	接続詞 副詞 文型 その他
上級後半	については/ときに まんざら~でもない/るわ~るわ にあって/べくもない/ようか/よもや	接続詞 文型 その他

4. データの提供方法

機能語用例データベース「はごろも」は、ダウンロード版と検索版の2つの形態で利用できる。ダウンロード版は Excel ファイルでリスト全体をダウンロードして使う形態、検索版は、ウェブインタフェースを使い、特定の項目を検索して使う形態である。

両者は想定ユーザーが異なる。ダウンロード版は、日本語教師が日本語教材などの教育コンテンツを作成する際の参考資料として使うことを想定しており、Excel ファイルもしくは CSV ファイルで運用している。一方、ダウンロード版は、日本語教師だけでなく、中上級者以上の日本語学習者が辞書的に引いて確認する使い方を想定しており、用例を表示することにフォーカスを置いている。4.1 節では、ダウンロード版について述べ、4.2 節では検索版について述べる。

4.1 ダウンロード版

機能語用例データベース「はごろも」のダウンロード版データは、<<http://hgrm.jp.org/>> に用意されたリンクから取得できる。データ取得時には、ウェブサイト上に用意された専用フォームを使って利用申請をする必要がある。利用申請と同時に、ダウンロード用のリンクが画面上に表示される。

ダウンロードファイルは zip ファイル形式になっており、zip ファイルを解凍すれば、CSV ファイル、XLS ファイル、XLSX ファイルの機能語用例データベース「はごろも」が出現する。いずれのファイルにも収録しているコンテンツは同じである。Microsoft 社の Excel プログラムを持っている利用者であれば、XLS ファイルか XLSX ファイルを、持っていない利用者であれば、CSV ファイルを選択する。

ファイルの項目は列 A から列 O までで、それらの情報を「収録情報」としてウェブサイト上に載せている (表 5)。

表 5: ダウンロード版データのテーブルフォーマット

列	項目名	説明	例, 注記
A	ID	項目の ID 番号を示す	1~1,884 (ただし, 欠番あり)
B	見出し	機能語項目 日本語教育において一般的に用いられている形式を用いている	V は動詞, A は形容詞 (イ形容詞), AN は形容動詞 (ナ形容詞), N は名詞, Vi は自動詞, Vt は他動詞
C	意味	見出し項目を特定するための特徴を示す意味・用法・機能・活用などを表す	例) 到着点, 逆接, 現在形 plain 肯定 (plain: 普通体, polite: 丁寧体)
D	見出し英訳	見出し項目の英訳を示す	例) at/in (Location)
E	意味英訳	意味の英訳を示す	例) Location of action
F	旧出題基準	旧出題基準の級を示す	4 級, 3 級, 2 級, 1 級
G	難易度	「はごろも」が独自に決定した 6 段階の困難度を示す	初級前半, 初級後半, 中級前半, 中級後半, 上級前半, 上級後半
H	例文	作例による典型例を示す	例) 公園でテニスをします。
I-O	参考資料のページ	見出し項目について, 7 種類の参考書のページを示す	『日本語表現文型』, 『現代語の助詞・助動詞』, 『日本語文型辞典』, 『現代語複合辞用例集』, 『辞書で引けない日本語文中表現』, 『日本語文法演習助詞』, 『複合助詞がこれでわかる』

列 A の「ID」は, 1,884 項目に対する固有番号である。列 B の「見出し」は, 日本語教育において一般的に用いられている形式を用いた文法項目の基本表記を示す。V は動詞, A は形容詞 (イ形容詞), AN は形容動詞 (ナ形容詞), N は名詞, Vi は自動詞, Vt は他動詞を意味する。次に, 列 C の「意味用法」には見出し語の意味機能に関する情報を示す。列 D の「見出し英訳」は, 見出し語の英訳を示す。列 E の「意味英訳」は, 日本語の「意味」を英訳したものである。列 F の「旧出題基準」には, 国際交流基金・日本国際教育協会 (2002) に準拠した日本語能力試験出題基準の文法レベルを示す。列 G の「難易度」は, 堀 (2012) で報告した情報に基づき, 初級前半~上級後半までの 6 段階の難易度が入っている。列 F の「例文」は, 項目の意味がよくわかるような例文を作例で示している。列 I から列 O の「参考資料のページ」は, 『日本語表現文型』, 『現代語の助詞・助動詞』, 『日本語文型辞典』, 『現代語複合辞用例集』, 『辞書で引けない日本語文中表現』, 『日本語文法演習助詞』, 『複合助詞がこれでわかる』の 7 つの出典から, 当該項目の記載があるページを示す。

4.2 検索版

機能語用例データベース「はごろも」は, コーパスから抽出した膨大な用例データを持っており, Excel シートによる運用には限界がある。また, 利用者によっては必要な項目だけ手っ取り早く確認したいという用途も考えられるであろう。これらを踏まえ, インターネットブラウザ上でデータを検索し, 用例を確認できるシステムを開発した。具体的には, 図 1 のような検索システムを開発した。

単語を入力して実行

しかし

部分一致 完全一致 検索 クリア

本システムについて 冠詞カテゴリーの一覧 検索結果

検索結果リスト

[「が」\(103件\)](#) [「しかし」\(103件\)](#) [「しかしながら」\(103件\)](#) [「もしも」\(103件\)](#) [「もしも〜ならば」\(103件\)](#)

(c) 2012-2015 「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」 科研グループ

図 1：部分一致によるキーワード検索の結果

図 1 の検索システムは、<<http://jreadability.net/hagoromo>> から利用できる。「部分一致」もしくは「完全一致」によるキーワード検索で必要な情報を取り出すことができる。デフォルトの設定では、部分一致による検索を行ったあと、「検索結果リスト」から項目を選定していく方法を想定している。図 1 は、「しかし」の部分一致による検索結果を表示している。部分一致をデフォルトとして設定したのは、項目見出しの表記に関するルールを知らなくても、文法項目の文字列の一部を入れることで必要な情報に到達できるようにする必要があると判断したからである。しかし、項目によって「検索結果リスト」が膨大になりすぎることもあるため、完全一致による検索オプションも用意している。例えば、「が」を部分一致で検索すると 103 件がヒットするが、完全一致で検索すると 2 件（格助詞の「が」と接続詞の「が」）がヒットする。

さて、見出し語が絞られた時点で、見出し語をクリックすると項目の詳細情報が表示される（図 2）。

見出し

しかし

意味	難易度	出典
話題転換	中級前半	辞書なし
意味	訳	
Topic conversion	However (to change a subject)	
例		
日本語文庫辞典：138ページ		
用例 (例)		
*1 すいぶん、ひどいね。しかしそれにしても佐田さん、遅いね。		
用例 (例)		
*1 F 1 3 2 : しかし、アメリカっていう国はちょっとず、すごい国だね。		
用例 (例)		
*0 しかし、メールが来ないって、さみしいよねー。		
用例 (例)		
*0 しかし、これでもらんと録音で済んだね。		
用例 (例)		
*0 (あーん) F 1 3 2 : あー、しかし、ほんと、だけだねー、あの、テレビな、テレビがね、24時間よ。		
用例 (例)		
*0 あー、しかしそりゃ結構だけど、悪いわけですよ。		

図 2：見出し語の表示

図 2 では、「話題転換」の用法の「しかし」の検索結果が表示されている。項目の意味や難易度、英訳、出典などの基本情報に加え、作例とコーパスの用例がそれぞれ表示されている。なお、Windows OS の場合、Google Chrome ブラウザと Opera に限定されるが、Google が提供する音声読み上げモジュールが組み込まれており、🔊 アイコンをクリックすれば、読み上げてくれる機能も搭載されている。Mac OS の場合、Google Chrome ブラウザのほかに、Safari、Opera でも読み上げ機能が利用できる。

リスト全体を俯瞰しながらデータ検索をする方法として、「意味カテゴリーの一覧」というタブが用意されている。このタブを選択すると機能語用例データベース「はごろも」リストの全意味カテゴリーを見ることができ、そこから見出し語を引くこともできる(図 3)。

- 不明確 (11)

かなんか	どうも	どこか	となく	とやら	なにかしら	なにやら	なんだか
なんとなく	やなんか	疑問詞～のやら					

- 不要 (4)

までのことはない	までのこともない	までもない	までもなく
----------	----------	-------	-------

- 並立 (8)

～も～なら～も～	ANで	Aくて	Nで	NやN	し	たり～たりする	とか
----------	-----	-----	----	-----	---	---------	----

図 3：意味カテゴリーの一覧

図3は、意味カテゴリーの具体例として「不明確」「不要」「並列」を表す文型の一覧を示している。意味カテゴリーからの検索は、書籍の索引でデータを検索していく方法に匹敵するもので、機能語検索に対する多様なニーズに答えていくために用意されたものである。

5. 今後の展開

今後は、より日本語教育にとって有効なデータベースとなるよう、次の4点を行ってきたい。

まず、データベースに格納する用例について、今後もコーパスからの抽出作業を続けて、公開する用例文を順次増やしていく。

次に、学習者のコーパスから抽出した例文を示す。例文は話し言葉、書き言葉の両方から収集し、文法的な正誤タグの付いているものを使用する。この目的は、利用者に正誤の確定した例文を示すことによって、①文法の形態的、意味的、語用論的な正誤を理解させること、②学習者が誤りやすい項目を理解させること、③学習者が当該項目についてどのような誤りを犯しやすいか理解させること、の3点が重要であると考えからである。

さらに、図3で示したように、現在意味カテゴリーには多くのものが未整理の状態に入っている。今後は分かりやすく整理し、意味索引としての利用が可能になるようにしたい。

最後に、ウェブページの閲覧者から記載項目に関するフィードバックを募り、項目の意味、難易度、用例などに関して知見を集め、よりよい検索システムにしていきたいと考えている。

*謝辞：本研究は、JSPS 科研費 22520538, 23242026, 15K02654 の助成を受けたものである。ことに JSPS 科研費 23242026 に関しては、「汎用的日本語学習辞書開発データベース構築とその基盤形成のための研究」(研究代表者：砂川有里子)の「機能語班」として加えられたことで、大きく発展させることができた。ここに記して謝意を述べたい。

参考資料

- 河原崎幹夫監修・木下直子他著 (1995) 『辞書で引けない日本語文中表現』北星堂書店
 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
 国立国語研究所 (1951) 『現代語の助詞・助動詞－用法と実例－』国立国語研究所
 国立国語研究所 (2001) 『現代語複合辞用例集』国立国語研究所
 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
 東京外国語大学留学生日本語教育センターグループ KANAME (代表鈴木智美) 編著
 (2007) 『複合助詞がこれでわかる』ひつじ書房
 中西久実子・庵功雄 (2010) 『日本語文法演習助詞』スリーエーネットワーク
 日本語記述文法研究会 (2009) 『現代日本語文法 第2巻』くろしお出版
 森田良行・松木正恵 (1989) 『日本語表現文型』アルク

文献

- 荒川みどり・木山 (2005) 「非母語話者日本語教師向け文法解説の試み」『国際交流基金日本語教育紀要』(1). 189-200. 国際交流基金
- 国際交流基金 (2012) 「2012 年度日本語教育機関調査結果概要抜粋」
<http://www.jpjf.go.jp/j/project/japanese/survey/result/dl/survey_2012/2012_s_excerpt_j.pdf>
- 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 (2012) 「日本語文法項目用例文データベース『はごろも』の概要, レベル分け, 文解析システムと, 利用可能性」パネルセッション「web ツールを通して世界とつながる日本語教育—文法用例文検索システム『はごろも』と国内外での利用可能性—」2012 年日本語教育国際研究大会
- 堀恵子・李在鎬・江田すみれ (2015a) 「文法項目用例検索データベース『はごろも』の開発と公開」第 19 回ヨーロッパ日本語教育シンポジウム
- 堀恵子・李在鎬・江田すみれ (2015b) 「機能語用例データベース『はごろも』の開発と公開」日本語教育学会
- 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
- 堀恵子・江田すみれ・李在鎬 (2010) 「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をめざして」2010 世界日本語教育大会 ICJLE2010) ポスター発表
- 堀恵子・江田すみれ (2011) 「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』の文法項目選定について」日本語教育学会 2011 年度春季大会
- 堀恵子・江田すみれ (2013) 「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」第 3 回コーパス日本語学ワークショップ 国立国語研究所

(2016 年 1 月 14 日受付)