

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月24日現在

機関番号：32663

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22520538

研究課題名（和文）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

研究課題名（英文）Development and Publication of the Database of grammatical items based on corpus with its searching tool

研究代表者

堀 恵子 (HORI KEIKO)

東洋大学・人間科学総合研究所・客員研究員

研究者番号：70420809

研究成果の概要（和文）：日本語教師支援のために文法項目の用例文を、話し言葉と書き言葉の複数のコーパスから抽出し、データベースを構築し、web上で公開する。文法項目は、日本語教育で教えられ、評価基準に用いられているものを1,884項目選んだ。文法項目に対し、日本語教育経験者による主観判定実験を行い、6段階レベルづけを行った。用例文は、5種の書き言葉コーパス、4種の話し言葉コーパスから抽出した。文解析システム「学習項目分析システム」と用例文データベースを公開する。

研究成果の概要（英文）：We developed the example sentences database extracted from two types of corpora, five written corpora and four spoken corpora. The system is opened to the public on the WEB site.

1,884 Grammatical items are chosen among items which have been taught and chosen as criterion of language tests. The items are classified into 6 levels from introductory level to advanced level by subjective evaluation participated by experienced Japanese language teachers.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	1,300,000	390,000	1,690,000
2012年度	600,000	180,000	780,000
年度			
年度			
総計	3,000,000	900,000	3,900,000

研究分野：人文学

科研費の分科・細目：日本語教育・言語学

キーワード：文法項目・データベース・難易度レベル・用例文・検索・web公開・教材開発・評価

1. 研究開始当初の背景

昨今コミュニケーションのための文法項目見直しが叫ばれている。これまで日本語学の知見に依拠していた文法項目選択ではなく、学習者がコミュニケーションを果たすため

に必要な文法項目を教えることの必要性が見直されている。

しかしながら、一方、日本語能力試験は改定され、新試験の出題基準は公開されておらず、これまでの出題基準に頼った文法項目の選

択とレベル分けは、今後はできない。また、旧出題基準には、実際には使用頻度の低い項目が含まれていることが複数の研究によって示されている。

そこで、教師自身が文法項目を十分理解し、それぞれの学習者のために必要な項目を取捨選択することの必要性が高まっている。そのためには、実際の用例に触れ、表現形態の違いによるニュアンスまで理解することが求められるが、海外にいる日本語教師、特に非母語話者の日本語教師にとってそれらは容易なことではない。

そのため、web 上で文法項目の用例が簡単に見るようなシステムを構築することができれば、日本語教師が文法項目の理解することに役立ち、海外の日本語教師支援となりうる。

2. 研究の目的

本研究の目的は、主に日本語教師支援として次の2点である。

(1) 文法項目の実際の用例に接することができるよう web を通して用例文を提供するシステムを構築する

(2) 教材開発、テスト開発等の支援ために文法項目の難易度を示す

3. 研究の方法

web システム構築の手順について5点に分けて述べる。

(1) コーパス収集

コーパスは話し言葉と書き言葉の次のを選んだ。

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)」京都大学情報科学研究科--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」(以下、BTS)

(2) 文法項目の決定 (先行文献)

文法項目はこれまで日本語教育で多く取り上げられてきた下記の5点の資料を参考に、基本的にはこのうちの2点以上に取り上げられているものを採用した。資料と採用の理由は、次のとおりである。

①「旧出題基準」: 初級から上級までの項目を網羅し、多くの教育現場、研究で参照されてきた。

②『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意味用法を解説している。

③『現代語の助詞・助動詞(以下、助詞・助動詞)』: 助詞、助動詞を扱っている。

④『日本語表現文型(以下、表現文型)』『現代語複合辞用例集(以下、複合辞)』: 複合辞を扱っている。

(3) 「はごろも」文法表作成

(2)の資料をもとに、文法項目を電子化し、意味用法、典型的な例文、先行資料を記入した一覧表を作成し、文法担当である江田・堀の合意の上で項目を選択した。

(4) データベース作成

「はごろも」は図1に示すように、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、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。

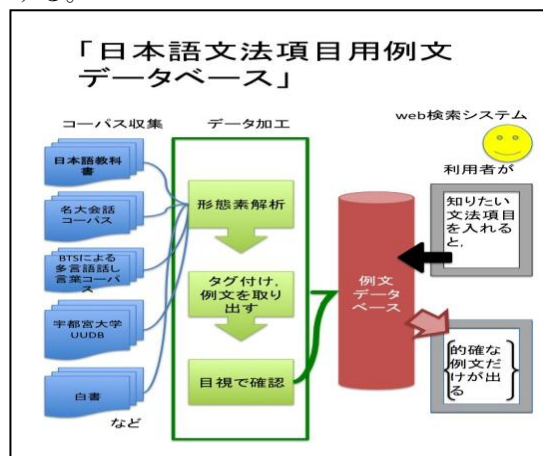


図1 「日本語文法項目用例文データベース」のイメージ図

(5)主観判定による難易度付け

文法表を教育現場での指導・教材作成・評価基準として使うためには、項目に難易度が付与されていることが必要である。そこで、主観判定による難易度付けを行った。

4. 研究成果

(1)「はごろも」文法表の作成

3. (3)に示したように、文法項目の一覧表を作成し、「はごろも」文法表と名付けた。また、用例文データベースのシステム全体を「はごろも」と呼ぶことにした。

文法項目は、1つの形式に複数の意味・用法、機能も取り上げた。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」だけが取り上げられていた。「はごろも」文法表では先行研究を参考に、下記の6つを取り上げることにした。

- 動作作用の継続「私は今本をよんでいます。」
- 結果の状態「まどがしまっています。」
- 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」
- 経験「彼は1ヶ月前に会社を辞めている」
- 恒常的な状態「道が曲がっている」
- 完了「子供が大学に入るころには、父親はもう定年退職しているだろう」

以上のような方法で、1,884項目を選んだ。

(2)文法項目6段階難易度付け

難易度は、CEFRの6段階を参考に6段階とし、「初級、初中級、中級、中上級、上級、超級」として、主観判定を行った。

調査協力者：日本語教師7名

判定方法：文法項目とその機能、例文を明示し、他の調査協力者の情報は示さないブラインド方式。

調査の結果：評定の一致度を示すκ統計量は、5名の評定者の間で0.53となり、中程度の一致が見られた。そこで5名の評定を採用し、その平均値をもって6段階レベルとした。レベルごとの項目数を表1に、また、「旧出題基準」とのクロス集計表とを表2に示す。

表1 「はごろも」6段階ごとの項目数

レベル	項目数	レベル	項目数	レベル	項目数
初級	155	中級	351	上級	523
初中級	196	中上級	592	超級	67

表2から、旧日本語能力試験級との関連について考察する。

「はごろも」文法表のうち、旧出題基準にある項目は955項目である。そこで、「はごろも」文法表の6段階レベル分けがどのよう

表2 「はごろも」6段階ごとの項目数

はごろも	旧出題基準				新しい項目	総計
	4	3	2	1		
6段階	4	3	2	1		
初級	127	6			22	155
初中級	54	59			83	196
中級	4	85	67		195	351
中上級		5	243	18	326	592
上級		1	82	119	321	523
超級			3	23	41	67
総計	185	156	395	160	988	1,884

なものであるかを推測するために、955項目について「はごろも」文法表のレベルと旧出題基準の級の順序性について分析を行った。

「はごろも」の6段階と旧出題基準の級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)を用いる。

その結果、γ=0.893となり、両者の間には高い相関が見られた。したがって、「はごろも」のレベル分けは、旧出題基準の級分けをかなり踏襲したものであると言える。

(3)「学習項目解析システム」の公開

学習項目解析システムは、<http://lias.intersec.tsukuba.ac.jp/>からアクセスでき、利用規約に同意すれば、無償で利用できる。



図2 学習解析システム

本システムは、生のテキストデータを自然言語処理の方法で下処理をしたあと、「はごろも」文法表の項目記述にマッチしたものを自動抽出する。例えば、「田中さんは学生です」に含まれる取立ての「は」および丁寧語の「です」、さらには「名詞+です」の文型を自動で抽出する(図2)。



図3 解析結果の表示画面

本システムを利用する利点として、1) 大量のデータに含まれる文法項目を簡単に取り出すことができること、2) 人手作業による漏れなどを防ぐことができることが挙げられる。これらの利点を生かし、コーパスに代表される生教材の定量的な評価に活用できると考えられる。

(4) 用例文データベースの公開

学習項目解析システムによって抽出された用例は目視の結果、用例文データベースが作成される。用例文は特に下位レベルにおいては膨大な数になるため、各コーパス5例を用例文として示す。出典のコーパスを明示することによって、表現形態による用例使用の傾向を示すことができる。

また、それ以外の用例については、意味用法を特定せず用例文群として提供する。それによって、文法研究に使用することも可能となる。

用例はweb上で順次公開される。

(5) 日本語教育のための文法研究

今回の活動は複数のコーパスを用いて条件にかなう例文を選択する作業を行った。その過程でいくつかの発見があり、それを日本語教育学会などで発表する機会を得た。日本語教育に貢献できる文法研究としては1) 「ている」「ていた」「ていない」に関するもの、2) 受身に関するもの、3) 「てくる」に関するものがあげられる。

「ている」については、「ていた」には「完了」「発見」など「ている」の知識だけでは捉えきれない用法があることを述べ、2013年に『「ている」「ていた」「ていない」のアスペクト—異なるジャンルのテキストにおける使用状況とその用法—』という本にまとめて発表した。

受身に関しては、会話とブログ・新書のコーパスを用い、受け身文の意味的な用法、文中の位置、文末表現、節との関係、ガ格の名詞、作用者格の名詞を調査した。

会話は「私」を主体とする直接受身、図書はモノ・コトの性質を、作用者を降格させて述べる、働きかけ性の薄い受身文が多かった。会話と書き言葉とでは文末表現が異なり、会話は述語のみの形は少数派で、補助動詞などの付いた形、従属節による言い終わりの形が多いのに対し、図書・ブログでは述語のみの形が多かった。作用者は3種のコーパス全部で表示されないことが多かった。会話では了解されていることが省略されるため、図書では能動主体を背景化させるためであった。

以上のように、受身は会話と図書で使われ方が大きく異なっている。この知見は初級でまとめて行われている受身の教育を、会話での受身、書き言葉での受身と分けて指導することにより効果をあげる可能性を示すものである。

「てくる」については、先行研究で行われている意味用法の下位分類のしかたが現実と異なるものがあることがわかった。移動の方法の用法が実際の用例ではよく見られるのに対し、そのことは文型辞典などでは触れられておらず、コーパスを使う研究の重要性がここでも示唆された。

(6) 今後の展開

日本語教育においては、文法項目のみならず、語彙項目が重要であることは言うまでもない。また、接続詞、文副詞などのように、語彙項目でありながら、文法、談話文法に深くかかわる項目も存在する。そこで、日本語教育に資する学習項目を包括的に示すためには、語彙研究とも連携していくことが必要となる。

現在、学習項目解析システムでは、文法項目のみならず語彙項目についても抽出して示している。このシステムの背景となっている学習辞書のための基礎研究や、他のweb上のシステムとの連携を深めていくことが必要である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

① 江田すみれ, テキストの違いと受け身文の違い—会話・ブログ・新書の受け身の使われ方をもとに—, テキストにおける語彙の分布と文章構造成果報告書 12-06, 無, 2013, 13-30

〔学会発表〕(計16件)

① 堀恵子・江田すみれ, 口頭能力評価に対して文法研究が貢献できること, 科学研究費助成事業合同成果発表会『言語能力評価の最前線—運用力の評価を目指して—』, 2013年03月28日, 桜美林大学

② 堀恵子・江田すみれ, 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』」のレベル分けと学習者コーパスの比較, 第3回コーパス日本語ワークショップ, ~2013年03月01日, 国立国語研究所

③堀恵子, 条件形式を含む前置き表現の習得について—学習者と母語話者のコーパス調査から—, 公開シンポジウム「習った表現・見落としている表現を見直す—文法、対照研究、語用論、習得の視点から—」, 2013年02月23日, 日本女子大学

④江田すみれ, 「ていた」の用法—「ている」の知識を応用するだけでは十分でない「ていた」の使い方について—, 公開シンポジウム「習った表現・見落としている表現を見直す—文法、対照研究、語用論、習得の視点から—」, 2013年02月23日, 日本女子大学

⑤堀恵子, 文法項目データベース「はごろも」の海外における利用可能性をさぐる—口頭能力評価と読解教材分析から—, 日本語学習辞書科研2012年度第1回全体研究集会, 2012年09月02日, 筑波大学

⑥堀恵子, 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, 2012年日本語教育国際研究大会名古屋2012 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」, 2012年08月17日, 名古屋大学

⑦江田すみれ, 旧能力試験3級項目の会話・ブログ・新書のデータでの使われ方について, 2012年日本語教育国際研究大会名古屋2012 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」, 2012年08月17日, 名古屋大学

⑧江田すみれ, 「ていた」のテキストの違いによる機能の違い, 2012年日本語教育国際研究大会名古屋2012 パネルセッション「日本語教育における「言語学」の可能性」, 2012年08月17日, 名古屋大学

⑨堀恵子・江田すみれ・李在鎬 (ほか2名4, 5番目), 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名古屋2012, 2012年08月17日, 名古屋大学

⑩堀恵子, 文法項目検索システム「はごろも」の文法リスト作成, 日本女子大学術交流主催, 『はごろも』研究会共催シンポジウム「web でつながる日本語教育」, 2012年3月3日, 日本女子大学

⑪江田すみれ, 『日本語文法項目用例文データベースはごろも』を使った文法研究の一例—テキストの違いによる受け身文の違い—, 日本女子大学術交流主催, 『はごろも』研究会共催シンポジウム「web でつながる日本語教育」, 2012年3月3日, 日本女子大学

⑫李在鎬, 日本語学習者作文コーパスについて,

日本女子大学術交流主催, 『はごろも』研究会共催シンポジウム「web でつながる日本語教育」, 2012年3月3日, 日本女子大学

⑬堀恵子, 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』の概要と目指すもの, 日本語学習辞書科研2011年度第1回全体研究集会, 2011年8月29日, 平泉文化遺産センター

⑭堀恵子, 文学・評論等書籍に現れた旧日本語能力試験1級文法項目の特徴—コーパス調査結果から—, 世界日本語教育研究大会 ICJLE2011, 2011年8月1日 天津外国語大学, 中国

⑮堀恵子・江田すみれ, 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, 2011年度日本語教育学会春季大会, 2011年5月21日

⑯堀恵子・江田すみれ・李在鎬, 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), 2010年8月1日, 国立政治大学, 台湾
〔図書〕(計1件)

⑰江田すみれ, くろしお出版, 「ている」「ていた」「ていない」のアスペクト—異なるジャンルのテキストにおける使用状況とその用法—, 2013年, 291ページ

〔その他〕

ホームページ等

学習項目解析システム

<<http://lias.intersec.tsukuba.ac.jp/>>

「日本語文法項目用例文データベース『はごろも』」プロジェクト

<http://jisho.jpn.org/?page_id=248>

6. 研究組織

(1) 研究代表者

堀 恵子 (HORI KEIKO)

研究者番号: 70420809

(2) 研究分担者

江田 すみれ (GODA SUMIRE)

日本女子大学・文学部・教授

研究者番号: 80205583

李 在鎬 (LEE JAE-HO)

筑波大学・人文社会科学研究所 (系)・准教授

研究者番号: 20450695

(3) 連携研究者 (0)