

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

堀 恵子
東洋大学人間科学総合研究所

キーワード：日本語教育 例文 コミュニケーション 日本語能力試験 機能語 複合辞

1. 研究の背景

言語教育の目的は、円滑なコミュニケーションができる能力を養うことである（木村ほか 1989：66）。Widdowson(1978)は、文法項目の選択に関して、コミュニケーションのためには学習者の目的に直結した言語使用の例として起こる可能性の高い項目を選ぶとしている。それを実現するためには、学習者の言語使用領域において、どの項目が実際に使用されやすいのかわかなければならないが、一般的な日本語教師が文法項目の頻度情報を手に入れることはそれほど容易なことではない。

また近年、日本語教育で扱われている文法の選択は日本語学に依拠してきた（野田 2005）と指摘され、見直しを迫られている。

これまで教育現場では、日本語能力試験に合格することが学習者の目標の一つであることから、教えるべき文法項目として旧日本語能力試験の出題基準（以下、「旧出題基準」）が参考にされてきた。しかし、堀ほか(2009)、堀(2009)は、旧出題基準1級99項目について頻度調査を行ったところ、使用例のない9項目や低頻度の項目があることを指摘している。さらに、日本語能力試験は改定され、新しい出題基準は今のところ非公開である。教育現場の教師、特に海外の非母語話者日本語教師からは何を教えるべきかわからないと戸惑いの声を聞く。

コミュニケーションに有用な文法項目を選択するためには、教育現場の教師が文法項目の意味用法、機能、頻度情報、使用される表現形態などについてよく理解し、学習者の目的にあっているかを判断することが求められる。そのためには、実際の用例を通してこれらを知ることは有益なことであると言える。

そこで、本研究ではweb上で文法項目の用例を検索して見ることができるシステムを開発しようとするものである。

2. 『日本語文法項目用例文データベース』の概要

2.1 システムの概要

『日本語文法項目用例文データベース』を端的に表すと、主に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、web上で公開しようとするものである。

2.2 想定される利用者

前章で述べたように、本システムは主に日本語教師を支援するものであり、想定される利用者としては国内・海外の日本語教師が挙げられる。特に実際の用例に触れる機会の少ない海外

の非母語話者日本語教師に有益な情報を提供できると考える。その他、自ら用例を読むことによって文法項目が理解できる上級以上の学習者や、日本語学、日本語文法に興味のある研究者、学生などの利用も想定される。

2. 3 システムの特徴

本システムの特徴として次の5点が挙げられる。

(1) 文法項目は初級～上級のレベルを網羅的に収集していること

助詞・助動詞・接続助詞・句型「たとえ～ても」・複合辞「～うが、まいが」などを含む。

(2) 構造化されたデータベースであること

用例の抽出は、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素材を収集する。その上で、意味・用法に一致するかどうかを目視で精選する。従って、単なる文字列検索とは異なり、的確でない例を含まない。

(3) 用例は、日本語教育の教科書、話し言葉、書き言葉のコーパスから抽出すること

書き言葉だけでなく、教科書、話し言葉のコーパスから抽出するため、表現形態、レジスターの違いによる使用頻度の偏りが理解できる。

(4) 1つの形式に複数の意味、機能も取り上げること

1つの形式の主要な意味・用法を取り上げるので、複数にわたることがある。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」だけが取り上げられていた。本システムではその他に、「完了」と「経験」の例も取り上げることにする。

(5) 文末表現に音調を付加（口頭表現）

話し言葉のコーパスからも用例を収集するため、音調が意味・用法の判別に関わる場合は、文末に音調情報を付加する。例えば、「でしょう」↗「でしょう」↘「でしよ」↗「でしよ」↘のように簡略化した記号で表すようにする。

2. 4 既存のツールとの比較

ここでは現在 web 上で使える既存の2点のツール、『現代日本語書き言葉均衡コーパス』(BCCWJ)の文字列検索「少納言」(以下、「少納言」と松吉ほか(2007)による「日本語機能表現辞書つつじ」(以下、「つつじ」と、本システムとを比較する。

2. 3 で示した特徴を持つため、「少納言」のような文字列検索においては的確でない表現形式が含まれるが、本システムでは意味に一致するものだけが取り出せることが優れている。

例えば、「でもいい」を調べたい場合、「てもよい」「てもよく」「てもよければ」という形式もあるため、「てもよ」の形で検索したとする。すると、「てもよかろう」「てもよさそう」のような的確な例と共に、「とてもよく油汚れを…」のような的確でない例も抽出されてしまう。本システムでは、目視によって不的確な文を削除しているため、このようなノイズが含まれることがない。

次に「つつじ」と比較すると、本システムは扱う文法項目が多いこと、用例文が示されることが挙げられる。「つつじ」は、助詞、助動詞、複合辞といった機能表現の意味、難易度、機能、文体などの多様な情報を提供するが、用例は直接的には示されていない。また、大見出しが項目341、2番目の意味による見出しも435項目であるのに対し、本システムでは大見出しが803項目である。本システムでは旧出題基準をほぼカバーし、旧1級、2級については、コ

一パスごとの頻度情報も提供する予定である。従って、頻度の高い項目を選んで教える時の判断材料とすることができる。

また、上記2ツールでは文末表現の音調は示していないが、意味・用法を判別するために重要な手がかりとなる場合があるため、2.3の(5)で示したように、話し言葉の場合は記号で示すことにする。

2.5 利用するコーパス

本システムが利用するコーパスは、下記の9点である。

(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/>

「白書」

「CASTEL/J CD-ROM V1.5」日本語教育支援システム研究会

「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

「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」

2.6 本システムのイメージ

以上から、本システムのイメージを図示したものが、次ページの図1である。

3. 文法項目について

3.1 文法項目の選定

文法項目を網羅的に扱っている先行研究には、次の5点が挙げられる。

- a. 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もの:「旧出題基準」
- b. 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もの:『日本語文型辞典(以下、文型辞典)』
- c. 助詞、助動詞を扱ったもの:『現代語の助詞・助動詞(以下、助詞・助動詞)』
- d. 複合辞を扱ったもの:『日本語表現文型(以下、表現文型)』、『現代語複合辞用例集(以下、用例集)』

本システムで取り上げる文法項目の決定において、これらの先行研究が広く日本語教育にお

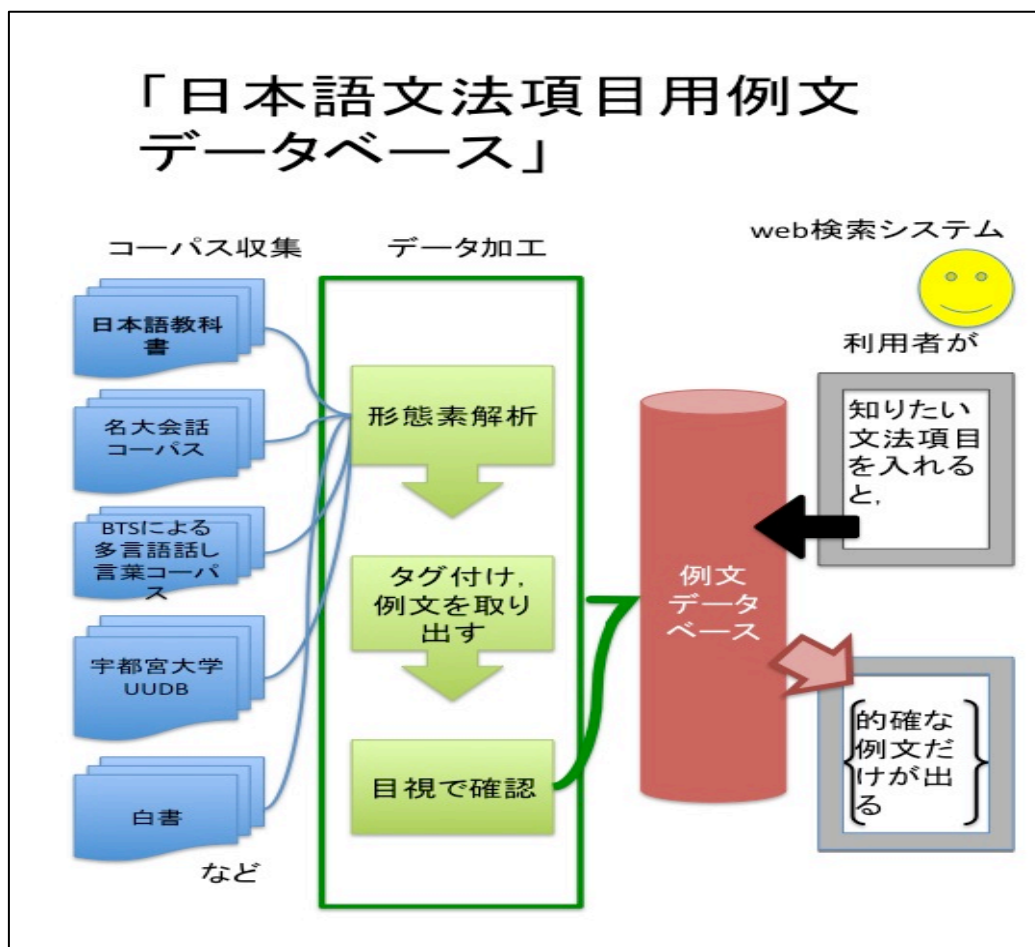


図1 「日本語文法項目用例文データベース」のイメージ図

いて用いられているため、参照することとした。

『用例集』には、複合辞に関して、表現文型、助詞・助動詞、文型辞典、用例集で取り上げている項目の一覧表がある。それに旧出題基準を加えた対照表を作成し、以下の2点を原則として項目を決定した。

- (1)旧出題基準の項目は、これまでの日本語教育に与えてきた影響を考慮し、活用事項などを除いて基本的にすべて採用する
 - (2)その他については、基本的に2点以上に記載があるものを取り上げる
- (2)の理由として、1点にのみ記載がある項目はわずかな用例のみしかない項目であると考えられるためである。

例) かとすれば、かとみると

以上、5点の先行研究を参考に、頻度や日本語教育的観点から大見出し803項目に決定した。これについては、今後の作業を通して精査することがありうる。

3. 2 文法項目の階層化

文法項目を具体的に表記する際には、活用、表記などの違いが問題となる。「つつじ」では、9段階の階層を設けているが、その中には「とりたて詞の挿入」「音韻変化」「です／ますの有

表1 表現文型、助詞・助動詞、文型辞典、用例集、旧出題基準の対照表(部分)

	「用例集」見出し形	表現文型	助詞・助動詞	文型辞典	「用例集」	(備考2)	旧出題基準
複合辞 001	あげく(に)	○	-	○	-		2
複合辞 002	あとで	-	-	○	○	* 1	4
複合辞 003	あとに	-	-	○	○	* 1	-
複合辞 004	以上(は)	○	-	○	-		2
複合辞 005	一方だ	○	-	○	○		2
複合辞 006	うえ(で)	○	-	○	○		2
複合辞 007	うえ(に)	○	-	○	○		2
複合辞 008	うえは	○	-	○	○		2
複合辞 009	得る	-	-	○	○	* 1	2
複合辞 010	お～ください	○	-	○	-		3

無」など、自然言語処理において機械に形式を認識させる必要から、文体、音韻の異形態を網羅すべくありうる場合を尽くすために非常に細かい階層立てとなっている。しかし、本システムは日本語教師支援を主な目的とし、文法項目を見るのは人であるため、とりたて詞、丁寧体形式の挿入などは、意味・用法の弁別に影響がないかぎり特に項目を立てなくてもよいと思われる。

そこで、5段階の階層を設け、「大見出し」「意味用法」「形式異同(とりたて詞の挿入、活用、助詞の交替など)」「音韻変化」「漢字表記」とした。

例として、「かぎり」「限り」「かぎりは」などの形式で用いられる複合辞と、音韻の異なる異形態を持つ「くらい」の場合の階層は、表2で示したようになる。

表2 「かぎり」「くらい」の項目の階層

A 大見出し	B 意味用法	C(形式の異同)	D(音韻変化)	E(漢字表記など)
かぎり	最高限度	かぎり	かぎり	かぎり
かぎり	最高限度	かぎり	かぎり	限り
かぎり	最高限度	かぎりは	かぎりは	かぎりは
かぎり	最高限度	かぎりは	かぎりは	限りは
かぎり	条件	かぎり	かぎり	かぎり
かぎり	条件	かぎり	かぎり	限り
かぎり	条件	かぎりは	かぎりは	かぎりは
かぎり	条件	かぎりは	かぎりは	限りは
かぎり	条件	ないかぎり	ないかぎり	ないかぎり
かぎり	条件	ないかぎり	ないかぎり	ない限り
くらい	程度	くらい	くらい	くらい
くらい	程度	くらい	ぐらい	ぐらい

「かぎり」の最高限度を表す用例としては a. と b. が、条件を表す用例としては c. と d. がある。

[最高限度]

- 途中、何故か見渡す<KEY>限り</KEY>田んぼで周りに人はおらず、かかしだけが数十体ぬっと立っているところに迷い込んだ時は、少々不安になった。(ブログコーパス)
- その人を直接知っている人が生きているときには、そういう人にも会い、私に可能な<KEY>かぎり</KEY>の材料をあつめます。(図書コーパス)

[条件]

- ということは、当分の間、携帯電話は自爆しない<KEY>限り</KEY>、比較的安全な端末とし

て、発展を続けるのかもしれません。(ブログコーパス)

- d. こう言うと、すぐ軍備拡張論者のレッテルを貼りつけられるであろうが、私としては、近代国家を信じている<KEY>かぎり</KEY>、軍備はどこかで必要であり、拡大する傾向からのがれられないということを論じたいだけである。(図書コーパス)

4. 結果の予測

4. 1 文法項目の頻度

堀ほか(2009)、堀(2009)は、旧出題基準1級99項目について、新聞、新書、論文、学会講演、模擬講演、改まった対話、くだけた対話の各コーパスを対象として頻度調査を行った(100万文字あたりの頻度を表示)。その結果、使用例のない9項目と使用頻度が低い項目があることが明らかになった。これについて堀(2009)は、旧出題基準を策定したときに使用したコーパスと、堀ほか(2009)、堀(2009)が対象としたコーパスとの違いが原因ではないかと述べている。そこで堀(2011)は、それを検証するために『現代日本語書き言葉均衡コーパス』モニター版公開データ(2009年度版)(以下、「BCCWJ2009モニター版」)の下位コーパスのうち、文学作品を対象として調査を行った。堀ほか(2009)、堀(2009)において、700万文字あたりの頻度が5以下の53項目について調査した結果、文学作品においても頻度が低い項目と、文学作品には頻度が高い項目とがあることが明らかになった。

使用例ゼロの語は「～ずにはすまない」「～ないではすまない」「ひとり～だけでなく/ひとり～のみならず」であり、頻度が0.2以下の語は「～ないものでもない」「～でなくてなんだろう」「～をものともせず」であった。

反対に高頻度の語は、「～なり」(例:「やめろ」 そう言うなり、テーブルをドンと叩いて立ち上がった。) 「～とばかりに」「～ばこそ」であった。

この結果から、本システムにおいても、同様に頻度の低い項目と高い項目のかたよりが見られることが予測される。旧1級と2級に関しては、コーパスごとの頻度情報を載せる予定である。表現形態、レジスターの違いによる頻度情報は、日本語教育への示唆を与えるものと期待される。

4. 2 日本語教育への示唆

本システムは、文法項目の用例文を提供し、文法項目に関する理解を深めることによって日本語教育に対して直接的な貢献ができるほか、表現形態、レジスターの違いによる頻度情報を与えることによって、学習者の目的に沿って教える文法項目を取捨選択ができるよう資料を提供することができる。

一般的にアカデミックな場面における日本語を学ぶ場合には、「新聞記事対応付けデータ」、 「CASTEL/J」などの書き言葉や、「日本語会話データベース」「名大会話コーパス」に見られる丁寧な話し言葉に使われる項目を教えることが有効であろう。また、日常生活のコミュニケーションに役立てるには、「ブログデータ」のような柔らかい文体の書き言葉や、会話コーパスの中のくだけた文体の会話に多く用いられる文法項目に注目することが有効であろう。また、日本語学、日本文学を学ぶ学習者には、4. 1で見たような文学作品に固有な項目を教えることが必要となる。

これまで旧出題基準1級項目はあまりコミュニケーションにおいて使われないと言われてき

たが、本システムが示すデータに基づいてどのコーパスにも使用例が少なければ、思い切って教えるのを止めることも、学習の効率をよくするために検討されるべきであろう。

このように、用例文を通して、日本語教育への示唆が与えられるものと期待する。

5. 今後の課題

システム開発の全体像は、図2の4つのステップを踏んで進むが、現在第3の段階を行っており、データ加工の済んだ用例文の候補を目視して用例文を挙げている。今後は、用例文が相当数上がった段階でweb上で順次公開をしていく予定である。

システムができあがったときには、日本語教育に関連した複数のwebツール（例えば、本プロジェクトである学習者辞書の項目や、リーディングちゅう太など）とのリンクができることで、学習者の利便性がより高まるであろう。

その場合の課題は、複数のシステム間の意味用法の関連づけである。各システムはそれぞれに、文法項目に意味・用法・機能のラベルをつけているが、利用者の利便性を考えると、どれかの意味・用法で検索することで、他のシステムが付けたラベルも参照できることが望ましい。

例えば、3.2で見た「かぎり」は本システムでは「最高限度」と「条件」の用法を立てている。しかし、下記のように先行研究によって与えるラベルは異なる。

用例集	「かぎりは」:	「順接・仮定」、「順接・因果関係」
文型辞典	「る/...ているかぎり」:	「範囲」
	「かぎり」:	「限界・極限」
	「ないかぎり」:	「不変化」

これらの関連づけをどう行うかが今後の課題である。

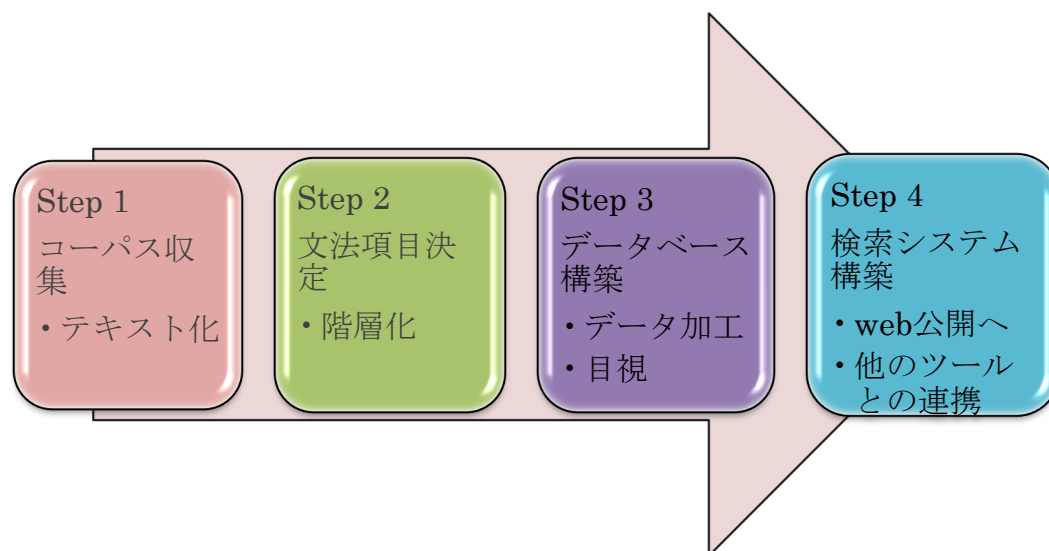


図2 システム開発の概念図

参考文献

Widdowson, H. G. (1978) Teaching Language As Communication. Oxford: Oxford Univ.Press

- (H.G.ウィドウソン(1991)『コミュニケーションのための言語教育』研究社出版)
- 木村宗男・阪田雪子・窪田富男・川本喬編 (1989)『日本語教授法』桜楓社
- 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
- 森田良行・松木正恵 (1989)『日本語表現文型』アルク
- 「日英新聞記事対応付けデータ (JENAAD)」Masao Utiyama and Hitoshi Isahara. (2003)
 Reliable Measures for Aligning Japanese-English News Articles and Sentences. ACL-2003,
 pp. 72--79.
- 「日本語機能表現辞書つつじ」<<http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji>> 名古屋大学 大学院工学研究科 電子情報システム専攻 佐藤研究室
- 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・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』,194-199.
- 堀恵子 (2009)「日本語能力試験文法出題基準の機能語を対象としたコーパス調査-表現形態と改まり度の違いに着目して-」2009年度豪州日本研究大会・日本語教育国際研究大会 (JSAA-ICJLE2009), 電子データ 151.
- 堀恵子(2011)「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-」2011年度日本語教育国際研究大会 (ICJLE2011) 予稿集『異文化コミュニケーションのための日本語教育』1, 855-856.