

令和 5 年 6 月 14 日現在

機関番号：12603

研究種目：基盤研究(C) (一般)

研究期間：2018～2022

課題番号：18K00708

研究課題名(和文)日本語教育語彙リストの開発ー読解語彙6000語の選定ー

研究課題名(英文)Creating a corpus based educational word list for learners of Japanese - 6000 words for reading -

研究代表者

本田 ゆかり (Honda, Yukari)

東京外国語大学・大学院国際日本学研究院・研究員

研究者番号：00817413

交付決定額(研究期間全体)：(直接経費) 1,000,000円

研究成果の概要(和文)：本研究は、日本語学習者が日本語を読む際に知っている役立つ「日本語読解基本語彙6000語」をコーパスと統計に基づいて選定し、語彙リストを作成したものである。リストの語彙は、日本語教科書に出現する語彙を調査して「初級日本語教科書出現語彙リスト」を作成し、日本語教育的観点から重要度ランクを調整した。また、漢字圏学習者と非漢字圏学習者では漢字語彙の学習しやすさが異なるため、日中同形語に関する情報を語彙リストの漢字語彙項目に付与した。また、本研究で作成した「初級日本語教科書出現語彙リスト」と「読解基本語彙6000語」を参照することができるWebツールを公開した。

研究成果の学術的意義や社会的意義

本研究は日本語テキストを読む際に役立つ基本語彙を6000語選定した。この語彙リストは日本語の書き言葉コーパスに基づき、汎用性の高い語彙を日本語教育的なレベル感を考慮して学習優先度順にランキングしている。そのため、この語彙は専門家の判定を主な基準として作られた語彙表の語彙と違うが、それは日本語の教師の直感が、語彙の使用実態や汎用性の高さと一致していないことを意味する。このように、コーパスと統計を基軸に語彙を選定し、量的研究に基づいて提示する総語数を決めてレベル分けを行うことによって、従来とは質的に異なる語彙リストを開発した点に本研究の学術的意義がある。

研究成果の概要(英文)：This study selected 6,000 basic words for Japanese reading comprehension, which are useful for learners of Japanese to know when reading Japanese, based on corpus and statistics, and created a vocabulary list. The vocabulary in the list is based on a survey of vocabulary appearing in Japanese language textbooks to create the "List of Vocabulary Appearing in Elementary Japanese Language Textbooks," and the importance ranks were adjusted from a Japanese language educational perspective. In addition, since the ease of learning Japanese Kanji vocabulary differs between Chinese L1 learners who know Chinese characters and others, information on Chinese-Japanese isomorphic words was assigned to Kanji vocabulary items in the vocabulary list. Moreover, a web tool that allows users to refer to the "List of Vocabulary Items Appearing in Elementary Japanese Language Textbooks" and "6000 Basic Reading Vocabulary" created in this study was made publicly available.

研究分野：日本語教育

キーワード：日本語教育 語彙リスト 基本語彙 コーパス 日本語教科書

1. 研究開始当初の背景

本研究はコーパスと統計に基づいて 6000 語の日本語教育語彙リストを作成するものである。教育語彙リストには二つの作成理念がある。一つは、「学習順序・学習段階に合わせて語彙を選ぶ」、もう一つは「将来的に知っておいたほうがよい語彙を言語使用の実態を見て選ぶ」という考え方である。前者のタイプはこれまで数多く作られてきたが、後者の例はまだ少ない。

前者のような語彙リストの開発は日本語教育の歴史が始まった頃から行われているが、その多くは比較的小規模な語彙調査の結果を資料として、日本語教育の専門家が経験に基づいて語彙を判断または判定したものである（以下、専門家判定方式）。その代表的な例に国際交流基金・日本国際教育支援協会編『日本語能力試験 出題基準』（1994）の語彙表がある（以下、「出題基準」）。このタイプは基本的に「学習順序・学習段階に合わせて語彙を選ぶ」という理念に基づいて作られている。しかし、どの語彙がどの学習段階に適切かという判断は専門家の主観によるため、このタイプのリスト間の語彙の一致率は低いという報告もある（饗庭 2011）。しかし、コーパスが言語研究資料として普及する前は、専門家判定方式が一般的であった。

これに対してコーパス準拠の語彙リストは、基本的に「言語使用の実態を見て選ぶ」という考え方による。その主なものに「日本語教育語彙表」（李・砂川 2011）、「日本語を読むための語彙データベース」（松下 2011）、「読解基本語彙 1 万語」（本田 2017）がある。

李・砂川（2011）の「日本語教育語彙表」は「現代日本語書き言葉均衡コーパス」（国立国語研究所 2011、以下、BCCWJ）に基づいて作成された語彙リストで、コーパスに基づいて選定された約 2 万語を収録している。また、この 2 万語には専門家の判定による学習難易度レベルも付与されている。このように「日本語教育語彙表」は、コーパス頻度に基づきつつも専門家の判定の比重も大きいので、上記の二つの理念においては折衷的な立場である。

松下（2011）「日本語を読むためのデータベース」は、約 6 万語の語彙を収録、うち 2 万語についてレベル分けをしている。元データは BCCWJ のうち書籍と Web テキストの 2 種類の媒体からなる約 3300 万語のコーパスであるため、語彙リストはこの 2 種の媒体の言語使用実態を反映したもので、多様な種類のテキストに対応した設計にはなっていない。また、3300 万語というコーパスは比較的小規模で、そこから 6 万語という多くの語彙を切り出してレベル分けを行うことは可能か、目的に対してコーパス規模は妥当か、という問題については議論の余地がある。

コーパスに基づいて教育語彙リストを作る際には、コーパスデータの収集が大きなポイントであり、語彙リストの使用目的に応じて標本抽出、代表制、均衡という三つの概念を考慮することになる（投野・本田 2016）。本田（2017）「読解基本語彙 1 万語」は BCCWJ（総語数約 1 億語）に基づくが、コーパスデータの収集に特に配慮した。このコーパスデータに基づき汎用性の高い語彙を選定しているため、専門家判定方式の語彙リストに含まれない項目も多数含まれる。そこで「読解基本語彙 1 万語」の評価のため、書籍、新聞、Web 上のテキスト、公開されている旧日本語能力試験の過去問題、話し言葉テキストを対象にテキストカバー率調査を行ったところ、旧日本語能力試験過去問題を含むすべてのテキストにおいて、「出題基準」よりも「読解

基本語彙 1 万語」ほうが高いテキストカバー率を示した。この結果は、データに基づく汎用性の高い語彙のほうが、専門家の直感によって選ばれた語彙よりも日本語能力試験を含む日本語のテキストを読む際に役立つ可能性を示唆している。

また、「読解基本語彙 1 万語」の作成過程で媒体に含まれる語彙の分布を調査したところ、コーパス出現頻度ランクの 6000 語付近に一つの閾値があり、それ以降の分布が不安定になることがわかった。すなわち、頻度ランク 6000 語程度までは媒体間で共通する語彙が多いが、この水準を超えると分野依存性が高くなり、語彙の共通性が顕著に低下する。これは語彙の汎用的な使用領域と、個別的な使用領域の峻別を示唆している。松下（2016）でも 7000 語付近に同様の傾向があることを指摘しているほか、英語教育では Tono（2013）もこの閾値について調査し、明らかにしている。このような語彙の使用領域の違いは、言語教育の観点からは基本語彙と専門語彙の違いとして解釈することができる。したがって、これらの先行研究を踏まえると、日本語教育基本語彙の語数は 6000～7000 語程度が妥当と考えられる。

2. 研究の目的

本研究は、本田（2017）「読解基本語彙 1 万語」精緻化し、6000 語のリストを作成することを目的とする（以下、「読解基本語彙 6000 語」。本田（2017）では「読解基本語彙 1 万語」も基本語彙としたが、厳密な意味で基本語彙と言えるのは上位 6000 語付近までである。それ以下の項目はコーパス・サンプリングによって大幅な順位移動が起こる可能性が高いため、日本語教育語彙として妥当なランキングを行うことは難しい。そこで、本研究では、汎用性の高い語彙を日本語教育的な重要度の側面から再検討し、日本語学習者に役立つ基本語彙リストを作成したい。

3. 研究の方法

6000 語は初級から中級で学習する範囲の語数である。そこで、「読解基本語彙 1 万語」の日本語教育的な学習優先度を再評価するため、以下 5 種の初級日本語教科書の出現語彙を調査する。

- ① スリーエーネットワーク（2012-2013）『みんなの日本語初級』
- ② 坂野永理，池田庸子，大野裕，品川恭子，渡嘉敷恭子（2020）『初級日本語げんき』
- ③ 嶋田和子監修（2011）『できる日本語 初級』
- ④ 国際交流基金（2013-2014）『まるごと入門、初級 1』
- ⑤ 東京外国語大学留学生日本語教育センター（2017）『大学生の日本語ともだち』

上記 5 種に出現する語彙を「初級日本語教科書出現語彙リスト」にまとめ、「読解基本語彙 1 万語」と統合して、教科書間の重なりが多い語彙から優先的に項目のランクを調整する。また、このリストをもとに、教科書語彙をチェックする Web ツールを開発し、公開する。

次に、日中同形語について調査する。日本語学習者人口の中で中国語母語話者は大きな割合を占めている。中国語と形が一致する日本語の漢字語彙には、意味も一致する同形同義語と、意味は異なる同形異義語がある。前者は中国語母語話者にとって学習上の負担が少ない。一方、後者のほうは学習負荷がかかるので、前者に比べて難しい項目であると言える。また、同形同義でも

意味範囲の重なりが完全一致ではなく部分的に一致するタイプは誤用を招きやすく、同形異義よりも習得が難しい場合がある。そこで、先行研究で示されている日中同形語の類似度を参照し、日本語教育経験のある中国語母語話者の協力を得て、中国語母語話者にとっての日本語の漢字語彙の難易度を検討する。

その他、日本語の教育や研究での使いやすさを考え、「読解基本語彙 1 万語」には入れていなかった項目（数詞、対概念語等）を 6000 語リストには入れる。また、「読解基本語彙 1 万語」では技術的な問題から一部の複合語が抜け落ちているので、日本語教科書コーパスと BCCWJ の長単位の頻度表に基づき、複合語を追加する。このように日本語教科書タグと日中同形語タグを付与し、「読解基本語彙 1 万語」で抜けていた項目をリストに追加したうえでランク調整を行い、日本語の学習や教育に配慮した「読解基本語彙 6000 語」を作成する。

4. 研究成果

4.1 初級日本語教科書に出現する語彙

前出の日本語教科書 5 種類に出現する語彙をリスト化し、「初級日本語教科書出現語彙リスト」を作成した。また、教科書間の語彙の重なりについても調査した。5 種類の教科書すべてに出現する語彙は、幅広い学習者が教科書という日本語テキストを読む際に役立つ語彙、すなわち読解基本語彙として比較的重要な語彙と見なすことができる。反対に、1 種類だけに出現する語彙は、その教科書の特徴的な語彙と考えることができる。この教科書語彙調査の結果を踏まえ、本研究の元データである「読解基本語彙 1 万語」に、重なる教科書数に合わせた日本語教科書タグを付与し、語彙項目のランク調整を行った。また、「読解基本語彙 1 万語」にない複合語も、「初級日本語教科書出現語彙リスト」を参照して 6000 語のリストに組み入れた。

4.2 「初級日本語教科書出現語彙チェッカー」の公開

上記の「初級日本語教科書出現語彙リスト」をもとに、
した (図 1, <https://textbooks.chuta.jp/>)。このツールは、調べたい単語やテキストに含まれる語彙が、調査対象とした 5 種類の日本語教科書のうち何種類に出現するかを
図 1 自動でリスト表示するもので、日本語教師がテストや教材で使う語彙を検討するのに使うことができる。

図 1 「初級日本語教科書出現語彙チェッカー」



4.3 日中同形語タグ付け

松下達彦・陳夢夏・王雪竹・陳林柯 (2017) 「日中対照漢字語データベース」 (Version 2.00) の「日本語独自の意味の類推可能性」という 5 段階評価を参照し、習得の難しさの点から漢字語彙の難易度を検討した。しかし、完全一致とそれに近い項目 (評価 5) 以外は、どのような順で易しいかということが判断しにくいという結論に至った。そこで、本研究では、語彙リストの利用しやすさを考慮し、研究協力者の協力を得て検討した中国語母語話者にとって習得が易しい

ことが明らかである評価 5 とそれに近い項目に対して、本研究の語彙リストにタグ付けした。

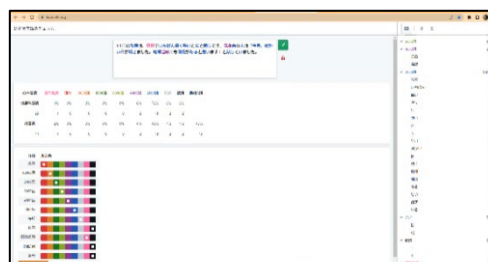
「読解基本語彙 6000 語」は日本語学習者にとって汎用性の面から学習優先度の高い語彙を選定したもののだが、中国語母語話者には学びやすい日本語の漢字語彙項目が少なからずある。本語彙リストを日本語の教育や研究に利用する際には、日中同形語タグを基準に学習優先度や、難易度について考えるとよい。また、今後は本研究で選定した「読解基本語彙 6000 語」のうち、中国語母語話者にとって学習しやすい日中同形同義語がどの程度含まれているかという問いについても、さらに調査を進めていきたい。

4.4 「読解基本語彙 6000 語」および「読解基本語彙チェッカー」の公開

4.1 の教科書出現語彙調査の結果を踏まえ「読解基本語彙 1 万語」の語彙のランク調整を行い、複合語と対概念語等を組み入れた「読解基本語彙 6000 語」リストを作成した。このリストの漢字語には、中国語母語話者にとって学習しやすい項目に日中同形語タグを付与した。現在、リストは最終確認と細かい修正作業を行っている。確認作業が終わり次第、Web 上に公開する（2023 年 10 月頃予定）。また、本研究では語彙リストを評価するためのテキストカバー率調査や既存の語彙リストとの比較等を行えなかったため今後の課題とする。

また、「読解基本語彙 6000 語」から語彙チェックツールを開発した（図 2, <https://basic.chuta.jp/>）。このツールは、調べたい単語やテキストに含まれる語彙が「読解基本語彙 6000 語」において汎用性および学習優先度がどのように示されるかを検索し、リスト化するものである。現在、Web 上で公開中のツールは「読解基本語彙 1 万語」のリストに準拠しているが、「読解基本語彙 6000 語」の語彙リストの最終確認および修正作業が終了した後（2023 年 10 月頃公開予定）、ツールに実装する予定である。

図 2 「読解基本語彙チェッカー」



<引用文献>

- ① 饗場淳子（2011）「日本語教育用語彙に共通する語についての一考察」『早稲田大学大学院教育学研究紀要』, 18-2.
- ② 国際交流基金・(財)日本語国際教育支援協会（2006）『日本語能力試験出題基準改定版』凡人社.
- ③ 投野由紀夫・本田ゆかり（2016）「第 2 章 教育語彙への応用」砂川有里子（編）『講座日本語コーパス 5 コーパスと日本語教育』朝倉書店, 35-57.
- ④ 本田ゆかり（2017）「コーパスに基づく『読解基本語彙 1 万語』の選定」, 『日本語教育』172 号, 118-132.
- ⑤ 李在鎬, 砂川有里子（2012）「コーパスを活用した日本語教育語彙表の構築」2012 年日本語教育国際研究大会パネルディスカッション 日本語教育につながるコーパス研究—現状と今後の展望—.
- ⑥ 松下達彦（2011）「日本語を読むため語彙データベース」（参照先：<http://www17408sui.sakura.ne.jp/tatsum/database.html>）
- ⑦ Tono, Y. (2013), Sampling biases and implications for better wordlist creation. Vocab @Vic conference, presentation slides. Victoria University of Wellington.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 本田ゆかり	4. 巻 172
2. 論文標題 コーパスに基づく「読解基本語彙1万語」の選定	5. 発行年 2019年
3. 雑誌名 日本語教育	6. 最初と最後の頁 118-133
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 本田ゆかり	4. 巻 172
2. 論文標題 コーパスに基づく「読解基本語彙1万語」の選定	5. 発行年 2019年
3. 雑誌名 日本語教育	6. 最初と最後の頁 118-133
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 本田ゆかり・川村よし子
2. 発表標題 「初級日本語教科書語彙チェッカー」の開発
3. 学会等名 カナダ日本語教育振興会年次大会CAJLE 2022（国際学会）
4. 発表年 2022年

1. 発表者名 本田ゆかり・川村よし子
2. 発表標題 読解基本語彙チェッカーの開発
3. 学会等名 日本語教育学会2019年度春季大会
4. 発表年 2019年

1. 発表者名 本田ゆかり
2. 発表標題 「初級日本語教科書共通語彙リスト」の開発
3. 学会等名 第52回日本語教育方法研究会
4. 発表年 2019年

1. 発表者名 本田ゆかり・川村よし子
2. 発表標題 読解基本語彙チェッカーの開発
3. 学会等名 日本語教育学会2019年度春季大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

読解基本語彙チェッカー https://basic.chuta.jp/ 初級教科書出現語彙チェッカー https://textbooks.chuta.jp/ 初級日本語教科書共通語彙リスト http://www.tufs.ac.jp/common/jlc/kyoten/development/shokyu.html

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------