

様 式 C - 19、F - 19 - 1、Z - 19（共通）

科学研究費助成事業

研究成果報告書



令和 6 年 4 月 5 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「現代日本語書き言葉均衡コーパス」（BCCWJ）に基づいて「汎用性を重視した日本語教育基本語彙6000語」（6000語リスト）を作成した。本田（2019）では頻度と分布統計によって1万語の語彙リスト（1万語リスト）を開発した。このリストをさらに使いやすくするため、日本語教科書の語彙調査を踏まえたランクの調整を行い、精緻化して「6000語リスト」とした。また、「6000語リスト」では「1万語リスト」の問題点も修正した。さらに、漢字語彙には漢字圏学習者にとっての語彙の難易度示す日中同型同義語タグを付与した。最後にこの語彙表をもとに、語彙レベルを判定する日本語学習支援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も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 developed a list of "6,000 Basic Japanese Vocabulary Words from a Versatility Perspective" ("6,000-word list") based on the Balanced Corpus of Contemporary Written Japanese (BCCWJ). Honda (2019) developed a 10,000-word vocabulary list ("10,000-word list") by means of frequency and distribution statistics. To improve this list, the ranks were adjusted and refined based on a vocabulary survey of Japanese language textbooks to create the "6,000-word list". The "6,000-word list" also corrects problems with the "10,000-word list". Furthermore, the kanji vocabulary was assigned a Japanese-Chinese isomorphic synonym tag indicating the difficulty level of the vocabulary for learners of kanji-speaking countries. Finally, based on these vocabulary lists, a Japanese language learning support web tool for judging vocabulary level was released.

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

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

語彙リストは言語教育においてさまざまな目的で作られ、利用されている。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」というもの、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、外国人に対する日本語教育の黎明期から行われてきた。その多くは、現在のコーパス研究と比較すると小規模な語彙調査の結果を資料として専門家の判断または判定によって語彙を選定したものである（以下、専門家判定方式）。その代表的な例に、『日本語能力試験 出題基準』（国際交流基金・日本国際教育支援協会編 1994）の語彙表がある（以下、「出題基準」）。このタイプの語彙リストは基本的に「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、専門家の主観によるものであるためか、リスト間の語彙の一致率が低いという報告もある（饗庭 2011）。コーパスが言語研究資料として一般的になる前は、専門家判定方式によって語彙リストが開発されていた。

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。このタイプの例としては、「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）、「日本語教育読解基本語彙 1 万語」（本田 2019）（以下、「1 万語リスト」）等がある。

現状では、「出題基準」のような従来型の語彙リストは数多くある一方で、コーパス準拠の語彙リストで信頼性の高いものは数が少ない。「入れたものが出てくる」というコーパスの性質を踏まえてコーパス・サンプリングを行い、コーパス規模に合う適切な語数を切り出し、日本語のテキストを読む際に知っておくと役立つ語彙を有用度順に配列した語彙リストがあれば、日本語の教育や研究に活用してもらえるのではないか。このような背景があり、本研究を実施するに至った。

2. 研究の目的

本研究は、本田（2019）の「1 万語リスト」に基づいて、「汎用性を重視した日本語教育基本語彙 6000 語」（以下、「6000 語リスト」）を作成することを目的とする。本研究では、コーパスと分布統計に基づき、日本語学習者が日本語のテキストを読む際に出現する確率が高い語彙、すなわち汎用性の高い語彙を選定し、日本語学習を効率的に行うのに役立つ基本語彙リストを作成したい。「1 万語リスト」の開発の過程で、コーパスの媒体（サブコーパス）に含まれる語彙の分布をクラスター分析したところ、コーパス出現頻度ランク 6000 語の付近に閾値があることがわかった。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。この閾値については、松下（2016）でも 7000 語程度の水準が示されているほか、Tono（2013）では英語コーパスを対象とした研究において同じ傾向があることを明らかにしている。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。

「6000 語リスト」では 「1 万語リスト」の問題を修正するとともに、日本語教科書語彙調査の結果も取り入れたい。教科書は初～中級の学習者が読む主要な日本語テキストである。そこで、現在使われている主な日本語教科書の語彙を調査して「初級日本語教科書出現語彙リスト」を作成し、「6000 語リスト」と統合して、項目のランクを調整する。

さらに、この「6000 語リスト」と「初級日本語教科書出現語彙リスト」をもとに Web ツールを開発する。学習や教育においてさまざまな語彙を扱うが、それらが重要語彙で用法なども含めて深く学習すべき語彙なのか、または話題やテーマに特有な語彙であり意味を知る程度でよいのか、という判断をするのは容易ではない。このような問題意識から、本研究で作成する「6000 語リスト」と「初級日本語教科書出現語彙リスト」に基づいて語彙チェックツールの開発も併せて行いたい。

3. 研究の方法

日本語教育では、6000 語は初級から中級までの学習範囲と考えられている。このレベルの学習者にとって、母語話者向けに書かれたテキストを自由に読むことは難しい。しかし教科書は、学習者のレベルに合わせて難易度がコントロールされていて、ほとんどの学習者が日々の学習を通じて読んでいる。このことから、教科書は初～中級の学習者にとって最も馴染みのある日本語テキストであると言える。本研究では、現在日本語教育でよく使われている以下 5 種類の初級教科書の出現語彙について調査し、日本語学習者にとって目にする確率が高い語彙をリストアップする。

- 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』

次に、日中同型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る。これらの学習者にとって、中国語と形や意味が似ているかどうか、日中同型語であるかどうかは語彙の難易度や習得のしやすさと深く関わっている。日中同型語には、形も意味も一致する同型同義語と、形は同じでも意味は異なる同型異義語がある。前者は中国語母語話者にとって学習しやすいが、後者は学習負荷が重く難しい項目である。また、同型同義でも意味の重なりが完全一致ではなく部分一致するタイプもある。これは誤用を招きやすく、同型異義よりも習得が難しい場合もある。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し、語彙リストにタグ付けを行う。

最後に、「1万語リスト」の問題点（数詞・対概念語、複合語、異表記等の項目の抜け）を修正する。これらの情報の追加、修正、ランクの再配列を行い、「6000語リスト」を作成する。

4. 研究成果

4.1 「汎用性を重視した日本語教育基本語彙 6000語」（「6000語リスト」）を開発した（表1）。「6000語リスト」は、日本語学習者にとっての語彙の汎用性を基軸に語彙を選定し、レベル別に分類したものである。日本語学習者向けにサブコーパスバランスを調整したコーパスの語彙頻度と分布統計に基づいて、全体としては約3万語を収録している。特に重要な上位6000語については汎用性レベルを2000語区切りに提示している。語彙リストの構成は以下の通りである。

- ① 順位：頻度×分布統計の値を軸に、単語親密度、および教科書における出現の有無によってランク調整をした語彙の汎用性のランク
- ② 汎用性レベル：1-2000、2001-4000、4001-6000、6001以上、の4段階
- ③ 見出し語：見出し語と追記情報
- ④ 読み：カタカナによる読み
- ⑤ 初級教科書：教科書における出現の有無
- ⑥ 日中同型語：中国語から特に類推しやすい項目に限りタグを付与
- ⑦ 出題基準：旧日本語能力試験の級
- ⑧ 行ラベル：品詞情報、他
- ⑨ コーパス出現頻度
- ⑩ DP：分布統計の値

表1 「汎用性を重視した日本語教育基本語彙 6000語」（「6000語リスト」）

順位	汎用性レベル	見出し語	読み	初級教科書	日中同型語	出題基準	行ラベル	総計（頻度）	DP	DP×対数
順位	汎用性レベル	見出し語	読み	初級教科書	日中同型語	出題基準	行ラベル	総計（頻度）	DP	DP×対数
1	1-2000	する	スル	◎		4	ある_スル_和_動詞-非自立可能_サ行変格	2282435	0.09	1
2	1-2000	なる	ナル	◎		2	成る_ナル_和_動詞-非自立可能_五段-ラ行	515664	0.12	2
3	1-2000	いる【～ている】	イル	◎		4	居る_イル_和_動詞-非自立可能_上一段-ア行	1031909	0.13	3
4	1-2000	ある	アル	◎		4	有る_アル_和_動詞-非自立可能_五段-ラ行	871014	0.17	4
5	1-2000	事	コト	◎		3	事_コト_和_名詞-普通名詞-一般_コト	665652	0.17	5
6	1-2000	出来る	デキル	◎		2	出来る_デキル_和_動詞-非自立可能_上一段-カ行	189132	0.18	6
7	1-2000	物	モノ	◎		4	物_モノ_和_名詞-普通名詞-サ変可能_モノ	253986	0.21	7
8	1-2000	中	ナカ	◎	◎	2	中_ナカ_和_名詞-普通名詞-副詞可能_ナカ	107152	0.20	8
9	1-2000	見る	ミル	◎		4	見る_ミル_和_動詞-非自立可能_上一段-マ行	204553	0.22	9
10	1-2000	持つ	モツ	◎		4	持つ_モツ_和_動詞-一般_五段-タ行	85552	0.21	10
11	1-2000	無い	ナイ	◎		4	無い_ナイ_和_形容詞-非自立可能_形容詞	427429	0.24	11
12	1-2000	取る	トル	◎		4	取る_トル_和_動詞-一般_五段-ラ行	74611	0.22	12
13	1-2000	来る	クル	◎		4	来る_クル_和_動詞-非自立可能_カ行変格	223351	0.24	13
14	1-2000	ため	タメ	△		3	為_タメ_和_名詞-普通名詞-副詞可能_タメ	130134	0.24	14
15	1-2000	行く	イク	◎		4	行く_イク_和_動詞-非自立可能_五段-カ行	200643	0.25	15
16	1-2000	上	ウエ	◎		4	上_ウエ_和_名詞-普通名詞-副詞可能_ウエ	47753	0.24	16
17	1-2000	前	マエ	◎		4	前_マエ_和_名詞-普通名詞-副詞可能_マエ	75448	0.25	17
18	1-2000	他	ホカ	◎	無し	2	他_ホカ_和_名詞-普通名詞-副詞可能_ホカ	37393	0.24	18
19	1-2000	所	トコロ	◎		4	所_トコロ_和_名詞-普通名詞-副詞可能_トコロ	99554	0.26	19
20	1-2000	言う	イウ	◎		4	言う_イウ_和_動詞-一般_五段-ワ行	710456	0.31	20
21	1-2000	内	ウチ			2	内_ウチ_和_名詞-普通名詞-副詞可能_ウチ	40767	0.25	21

まず、「6000語リスト」には、「1万語リスト」で除外していた対概念語や数詞等を戻し入れた。

そのうえで、初級日本語教科書語彙調査の結果をもとに、教科書によく出る語彙の項目ランクを上げるなどして調整し、2000 語区切りに「汎用性レベル」を定めた。初級教科書における出現は、◎が 5～4 冊に出現する語彙、○が 3～2 冊に出現する語彙、△が 1 冊に出現する語彙というルールで示されている。

さらに、「6000 語リスト」の漢字項目について、中国語母語話者にとって特に類推が容易な日中同型同義語に限定してタグを付与した。これは松下他（2017）の分類をもとに、日本語教育を専門とする中国語母語話者の協力を得て検討した。◎は松下他（2017）の評価のうち 5（とても類推しやすい）に当てはまるもの、○は 4（類推しやすい）に相当すると考えられるものである。これらの一部には、本研究で再検討して評価を変えているものもある。

4.2 Web ツールの公開

「初級日本語教科書出現語彙リスト」をもとに語彙検索ツールを開発し公開した（<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自動でリスト表示するものである。



図1 「初級教科書出現語彙チェッカー」出力画面

また、「読解基本語彙チェッカー」の開発も行った（<https://basic.chuta.jp/>）。このツールは、「6000 語リスト」に基づいて語彙の汎用性レベルをチェックするものである。現在公開中のツールは「1 万語リスト」に準拠しているが、近々、語彙リストを差し替える予定である。また、今後は、語彙リスト本体を Web サイトからダウンロードができるようにする計画である。

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
---------	---------