

国立国語研究所学術情報リポジトリ

日本語教育のための初級語彙をどう定めるべきか

メタデータ	<p>言語: Japanese</p> <p>出版者: 国立国語研究所</p> <p>公開日: 2025-07-25</p> <p>キーワード (Ja): 初級語彙リスト, 頻度, 使用範囲, 語彙体系, 語彙モジュール</p> <p>キーワード (En): elementary vocabulary list, frequency, range, lexical system, vocabulary module</p> <p>作成者: 松下, 達彦, 岩下, 智彦, 新井, 智大, ABUELLIL, Nora, 田中, 祐輔, 中俣, 尚己, スルダノヴィッチ, イレーナ, 松田, 真希子, 劉, 瑞利, 陳, 夢夏, 柏野, 和佳子, 石黒, 圭, 橋本, 直幸, 小熊, 利江</p> <p>メールアドレス:</p> <p>所属: 国立国語研究所, 電気通信大学大学院 博士後期課程, 明治大学大学院 博士後期課程, 早稲田大学大学院 博士後期課程, 筑波大学, 大阪大学, ユライ・ドブリラ大学プーラ, 東京都立大学, 中山大學, 寧波大学, 国立国語研究所, 国立国語研究所, 福岡女子大学, ルーヴァン・カトリック大学</p>
URL	https://doi.org/10.15084/0002000522

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



日本語教育のための初級語彙をどう定めるべきか

松下達彦^a 岩下智彦^b 新井智大^c ABUELLIL Nora^d 田中祐輔^e
中俣尚己^f スルダノヴィッチ イレーナ^g 松田真希子^h 劉 瑞利ⁱ
陳 夢夏^j 柏野和佳子^a 石黒 圭^a 橋本直幸^k 小熊利江^l

^a 国立国語研究所／総合研究大学院大学

^b 電気通信大学大学院 博士後期課程／国立国語研究所 非常勤研究員 / 共同研究員

^c 明治大学大学院 博士後期課程／国立国語研究所 非常勤研究員 / 共同研究員

^d 早稲田大学大学院 博士後期課程／国立国語研究所 共同研究員

^e 筑波大学／国立国語研究所 共同研究員

^f 大阪大学／国立国語研究所 共同研究員

^g ユライ・ドブリラ大学プーラ／国立国語研究所 共同研究員

^h 東京都立大学／国立国語研究所 共同研究員

ⁱ 中山大学／国立国語研究所 共同研究員

^j 寧波大学／国立国語研究所 共同研究員

^k 福岡女子大学／国立国語研究所 共同研究員

^l ルーヴァン・カトリック大学／国立国語研究所 共同研究員

要旨

初級語彙をリストとして示すことには一定の意義があり、現実には様々な語彙リストが教材作成、語彙テスト等に応用されている。しかし、現実にもよく参照される旧日本語能力試験の4級、3級の語彙表は社会の変化に適合しておらず、初級語彙リストを再検討する必要がある。本稿は、初級語彙リストのあり方を展望するため、まず、従来の各種リストの問題点を考察し、英語の語彙リスト研究を含めて基本語彙選定をめぐる議論や成果を概観した。そのうえで、初級語彙の基本的条件として頻度、範囲、語の体系とまとまりなどの7項目を挙げ、初級から日常生活、学術、就労・ビジネス、趣味・文化・芸術などの目的に応じた領域別語彙モジュールを作成すべきであることを主張した。さらに、初級語彙選定に利用できる語彙資料を挙げ、コーパスの語彙頻度の性質の評価、語彙頻度への重みづけ、分散度も考慮に入れた基準順位の決定、語彙表の語彙の頻度範囲の確認などの、11のステップによる初級語彙選定およびその妥当性検証の手順を示した。あわせて、作成上の留意点として、海外学習者のニーズへの配慮、場面と話題の視点、重要度と難易度の区別、受容と産出の区別、多義語の問題などを挙げ、どうあるべきかを整理した*。

キーワード：初級語彙リスト、頻度、使用範囲、語彙体系、語彙モジュール

* 本研究は、国立国語研究所の共同研究プロジェクト「多様な語彙資源を統合した研究活用基盤の共創」（プロジェクトリーダー：小木曾智信）のサブプロジェクト「学習者辞書用語彙資源の構築」（プロジェクトリーダー：柏野和佳子）及び、JSPS 科研費 JP 23H00072（代表者：松下達彦）の研究成果である。なお本稿は、2024年11月24日「学習者辞書用語彙資源の構築」共同研究発表会における口頭発表「日本語教育のための初級語彙をどう定めるべきか」をもとに加筆修正したものである。

1. 研究の背景

1.1 初級語彙リストの意義

一般に外国語学習では初級、中級、上級といったレベル分けがなされ、日本語教育においても初級語彙、初級文法と言え、理解に多少の幅はあるが、その範囲や分量においてある程度の共通理解があると思われる。例えば、「初級段階では基本的な文法・文型とともに、基本的な語彙約 1,500 ～ 2,000 語、漢字約 500 を習得させることを目安とし、その消化のために 250 ～ 300 時間が必要であると考えられてきた」(木村ほか 1989: 108), 「一般に初級段階で与える語彙数は 1,000 から 1,500 語ぐらい、多くても 2,000 語どまりであろう」(森田 1986: 98-99), 「一般に日本語教育の場で初級というのは、基本的な文型、語彙約 2,000 語、片仮名、平仮名、漢字 300 ～ 400 字、学習時間 300 時間を経た段階である」(北條 1989: 238) など、日本語教育実践から導き出された見解として語彙の数や学習時間について論じられている。また、田中 (2016) のように、複数の教材に共通して扱われる語彙を分析し、初級語彙を帰納的に導く取り組みもある。

質的に見ると、初級語彙には領域を問わず幅広く用いられる高頻度語彙という特徴がある。ほかにも考慮すべき点 (3.2 および 4.1 で後述) があるが、領域を問わず幅広く用いられる高頻度語彙であるという点は、初級語彙リストにほぼ共通していると思われる。

近年はヨーロッパ言語共通参照枠 (CEFR: Common European Framework of Reference for Languages) の影響を受け、「JF 日本語教育スタンダード」(国際交流基金) や「日本語教育の参照枠」(文化審議会国語分科会 2021) 等、言語能力記述 (Can-do statements) によるレベル表示が日本語教育界でされるようになってきている。初級は CEFR および「JF 日本語教育スタンダード」「日本語教育の参照枠」でいえば A1 から A2 程度までに相当する。CEFR では、A2 レベルについて「●ごく基本的な個人的情報や家族情報、買い物、近所、仕事など、直接的関係がある領域に関する、よく使われる文や表現が理解できる。●簡単で日常的な範囲なら、身近で日常の事柄についての情報交換に応ずることができる。●自分の背景や身の回りの状況や、直接的な必要性のある領域の事柄を簡単な言葉で説明できる。」(吉島・大橋 (訳・編) 2004: 25) としている。これを要約すれば、初級とは、日常生活領域で「よく使われる」(すなわち高頻度の) 言語の運用が可能なレベルということになろう。CEFR では B1 で個人的関心に、B2 レベルで専門分野に言及されており、初級を終えて中上級へと進むにしたがって、幅広い領域から、各自が目的とする特定領域の語彙に特化していくのが自然な流れだと考えられる。「JF 日本語教育スタンダード」も「日本語教育の参照枠」も CEFR のレベル設定とレベル記述をほぼ全面的に踏襲している。また、英語教育において多くの辞書や学習書を出版している Cambridge University Press では、English Vocabulary Profile として CEFR の A1 ～ C2 の各レベルに対応する語彙リストを作成しており、イギリス英語では、A1 で 784 語、A2 で 1,594 語 (計 2,378 語) を、アメリカ英語では A1 で 782 語、A2 で 1,564 語 (計 2,346 語) をリストしている (Cambridge University Press & Assessment 2024)。

初級とされている語彙 (例えば旧日本語能力検定試験の 4 級、3 級の文字・語彙の出題範囲として示された語彙) が、日常会話のみならず、新聞、文芸作品、学術テキスト等においても高頻度で用いられることは、テキストカバー率の研究によって確認されている (Matsushita 2012:

296-329, 松下 2016: 66-71, Matsushita 2023)。したがって, CEFR や「JF 日本語教育スタンダード」「日本語教育の参照枠」の A2 レベルまでに必要な語彙は, 一般的に初級語彙として理解されているように「領域を問わず幅広く用いられる高頻度語彙」とおおむね一致すると考えてよいであろう。(ただし, 4.2 で後述するように, 日常生活領域のみとすることが, 非日本語圏における初級学習者のニーズに合致するかどうか, 検討の余地があるものと思われる。)このような参照枠に基づいたレベル表示の共有が進めば, 複数機関のカリキュラムの比較や接続が容易になり, 学習者の移動に伴う問題も解決しやすくなるはずである。

レベルに応じた語彙が定まることにはいくつか利点がある。第一に, 教科書や教材の開発の基礎資料となる。特に初級語彙リストは重要である。すべての学習者が一度は通る学習段階であり, 初級で学習をやめる学習者が少なくないため, 中級・上級以上に影響が大きいと言えよう。また, 「やさしい日本語」の使用が地方自治体や観光, 商業の現場で広がりつつある今, 初級の語彙リストは現実の言語使用の指針にもなり得る。第二に, ある一まとまりの文章や談話に含まれる初級語彙の割合を測定することで, 語彙的難度を決める指標になる(岩田ほか 2015a)という点でも初級語彙リストは重要である。このような語彙的難度の測定ツール「語彙頻度プロファイラー」は, 日本語では, 「リーディング・チュウ太」(川村ほか), 「J-LEX」(菅長・松下 2013), 「jReadability¹」(Lee and Hasebe 2013), 「やさしにチェッカー」(岩田ほか 2015b), 「読解基本語彙チェッカー」(本田)²などがあり, 文章の語彙レベルの判定やリライトによる語彙レベルの調整に使われている。第三に, 学習者の語彙知識の測定のためのテストを作成する際には, 項目のサンプリングの基礎データになり, 文章の難度の測定結果と合わせてみることで, あるテキストが学習素材として学習者レベルに適合するかどうかを判定できる。

1.2 従来の初級語彙選定の問題点

ところが, 初級教科書の語彙を比較した先行研究を見ると, 教科書間の語彙の共通性は決して高くない(今西・神崎 2008, 田中 2016)。その原因としては, 初級では文法項目の扱いが中心であることや, 目的や対象により異なる語彙ニーズが存在する可能性が考えられる。1.1 で, 初級語彙は幅広い領域で用いられると述べたが, 初級教科書間の語彙の共通性は高くないのである。

そもそも日本語教育界では, 第二次世界大戦後は長らく旧日本語能力試験の 4 級, 3 級³の文字・語彙の出題範囲として示された語彙表以外に, 広く共通に参照される初級語彙リストが存在しない⁴。伊東(2006)は, 『日本語能力試験出題基準』について, 日本語教育の全般的な指針となっ

¹ 多岐にわたる機能の一部に語彙頻度プロファイラーが含まれている。

² 「読解基本語彙チェッカー」のサイトには制作者名がないが, 同サイトの「資料ダウンロード」から参照される資料は本田ゆかり氏によるものであり, 本田(2019)に記された <http://yukari.overworks.jp/> からでも当サイトに行きつくため, 同氏によるものと考えられる。

³ 日本語能力試験の主催団体は, 旧試験 4 級は現行試験 N5 レベルに相当し, 3 級は現行試験 N4 レベルに相当すると説明している(国際交流基金・日本国際教育支援協会)。

⁴ 戦中期においては『日本語基本語彙』(国際文化振興会 1944)などいくつかの語彙表が直接, 間接に参照されていた。また, 現在, 「日本語教育語彙表」(砂川ほか)も初級語彙を確定できる語彙リストであるが, 初級語彙のリストとして広く認知・利用されているとは言えないと思われる。

いて教育内容やカリキュラムの策定に少なからず影響を与えており、試験そのものがある種のスタンダードとしての機能を有することを指摘している。

最も普及した前出の旧日本語能力試験出題基準が発表される前には『日本語教育のための基本語彙調査』（国立国語研究所 1984）が公刊され、基本語彙に関する 7 種対照表に基づき「基本語二千」「基本語六千」が選定されていた。しかしながら、この 7 種の基本語彙表のうち、頻度データを含むのは、当時としては大規模な語彙頻度調査であった『現代雑誌九十種の用語用字』（語彙表）（国立国語研究所 1962）のみであった。しかも、旧日本語能力試験の 4 級、3 級の出題範囲の選定にあたって、1 次資料として参考にされたのは 11 種の教科書⁵の採録語彙であり、『日本語教育のための基本語彙調査』は 2 次的な資料として参照されただけであった。4 級については「11 種の教科書のうち 4 種類以上の教科書の 4 級学習範囲に共通して提出されている語彙」を候補語彙として、「使用領域、意味上の相互関係を考慮し、日本語教育に関する語彙の調査資料を参考にして」「認定を行った」（国際交流基金・日本国際教育協会 1994: 12）とある。そして、この出題範囲の語彙表は、その後の教科書や試験対策本を含む学習教材の作成に多大な影響を与えてきた。つまり、当時は、語彙頻度などの客観資料に基づいて教科書ができたのではなく、教科書が語彙のレベルを確定させる資料として参照されていたのである。確かに、初級について言えば、当時は日常会話の語彙の計数に使えるよい会話コーパスがなかったのも、初級語彙の選定に頻度を参考にすることは難しかったと言えよう。しかしながら、教科書を参照して選ばれた語彙が、その後の教科書の編集のために参照されれば、妥当性のレベルは初めに参照された教科書次第ということになり、データの循環参照となるため、客観性を担保できない。

加えて、旧日本語能力試験の 4 級、3 級の語彙表は、今の時代に合わない面も多い。

- ・社会の変化で低頻度となった語がある（「マッチ」「カセット」「字引き」など）
- ・社会の変化で高頻度となった語、特に外来語が少ない（「メール」「スマホ」など）
- ・治安関連の語「おまわりさん」「交番」「警官」「すり」が頻度レベル以上に収録されているが、「消防」「救急」はない。
- ・日本語能力試験開始当時は、留学生の入学選抜の目的で多く使用されていたため、大学以外では初級と言えない語（「研究」「輸出」「輸入」など）が採録されている
- ・オノマトペや野菜・果物の名前などが少ない。

また、重要度と難易度の混乱もある。「よく使われる」（＝重要度が高い）ことと「学習しやすい」（＝難度が低い）ことはイコールではない。初級語彙の確定には重要度も難易度も考慮する必要はあるが、難しいというだけで重要度の高い語を初級から外したり、学習しやすいというだけで初級に加えたりしてもよいだろうか。例えば「表示」という語の頻度順位は「現代日本語書き言葉均衡コーパス」の短単位語彙表（Version 1.1）（国立国語研究所）では 612 位で、「日本語

⁵ 当時の国内および海外における使用教科書調査に基づき、使用機関数の多い順に教科書を選び、収録語彙を調査したとある（国際交流基金・日本国際教育協会 1994: 3）が、具体的な教科書名は記されていない。

を読むための語彙データベース」(松下 2011a) の「書きことば重要度ランク」でも 1,086 位だが、旧日本語能力試験では級外語彙である。一方、「花瓶」という語は、旧日本語能力試験では 4 級(初級前半、現在の N5 相当)とされており、意味理解も難しくないと思われるが、頻度は 13,767 位(国立国語研究所)、11,772 位(松下 2011a)である。初級語彙を定めるうえで、どのような順序で重要度と難易度を考慮するのかを再検討する必要がある⁶。

2. 問題の所在と研究課題

以上をまとめると、これまでの日本語の初級語彙リストは、教師の経験や勘に基づく主観選定や、日本語教科書の採録語彙の重なるの調査によるものが大半である。最も影響の大きい旧日本語能力試験の 4 級、3 級の出題範囲の語彙の選定でも主観に基づく教科書の採録語彙が主な資料とされており、頻度が参照されたのは選定の 2 次資料として使われた『日本語教育のための基本語彙調査』の対照表の一部だけで、それも書き言葉の語彙頻度のみであり、時代に合わない部分や、重要度と難易度の混乱も見られるということである。

教科書語彙の選定の主観が「初級」であるべき語彙を正しく反映する保証はない。本質的に妥当性を欠く恐れのある初級語彙が影響力を持つことは問題である。本来は、明確な原理に基づく選定基準があり、それに応じたデータ分析に基づいて選定されるべきであり、科学的に妥当かつ信頼できる語彙リストに基づき、カリキュラムや教材が作成されることが望ましい。これまでは、基本語彙をめぐる議論(3.2 で後述)はあったものの、学習者ニーズや学習難易度を考慮した議論が乏しく、選定のデータも現在より乏しかった。現在でも基礎データは十分ではないが、旧日本語能力試験出題基準の開発当時に比べれば、日常会話の語彙表(国立国語研究所 2022)が作成されるなど、基礎資料が増えている。

そこで本稿では、初級語彙選定の原理と選定の基準や方法を検討する。これにより、例えば、既存の初級教科書の収録語彙を評価できるようになるであろう。初級語彙選定に関わる先行研究を踏まえ、以下の 3 点について論じ、初級語彙リストのあり方を展望する。

- 1) 「初級語彙」の選定基準はどうあるべきか。
- 2) 日本語学習者のニーズをどのように一般化し、目的に応じた何種類の初級語彙があればよいと考えるか。
- 3) 「初級語彙」選定の具体的方法はどうあるべきか

なお本稿では基本的に成人学習者用の語彙リストについて論じる。年少者に関しては、日本語

⁶ 単語親密度(word familiarity)を選定の基準として考える議論もある(佐藤ほか 2004, 獅々見 2016)が、親密度は主観であり、それが重要度(頻度が表す社会的ニーズ)を表すのか、難易度を表すのか、意味的な親近性(日常生活でよく見かける等)を表すのかが不明である。本研究では、難易度や意味的な親近性に関わらず、重要度を重視する立場をとるため、親密度は考慮しない。また、同じ語でも文字提示と音声提示と文字・音声の同時提示では親密度が大きく異なる(天野・近藤 2000)ため、親密度の利用は非常に難しいと考える。ただし、獅々見(2016)は基本語選定の対象がオノマトペであり、語彙頻度を親密度と併用する選定方法には一定の妥当性があると思われる。

母語の児童・生徒も含む問題になり、ニーズも異なるため、稿を改めて論じたい。

3. 先行研究

3.1 既存の初級語彙

ここでは原則として、成人の第二言語学習者を対象とする初級語彙リストを取り上げる。

主な日本語の初級語彙リスト、または初級語彙を特定できる語彙データには以下がある。(どこまでが初級かを示していない頻度データや、入手の難しいデータは除く⁷⁾。

- 1) 『日本語教育のための基本語彙調査』の「基本語二千」(国立国語研究所 1984)
- 2) 旧日本語能力試験の4級, 3級出題範囲(国際交流基金・日本国際教育協会 1994)
- 3) 「日本語を読むための語彙データベース」の初級(松下 2011a)
- 4) 「日本語教育語彙表」の初級(砂川ほか)

既に述べてきたように、1) を参照しつつ教科書データ及び主観を交えて2) が選定されているのであるが、そこには多くの問題がある。3) には3種類の語彙ランキングがあるが、「留学生用」「一般用」のランキングでは、初級に関しては2) の語彙表を参照している。「書きことば」ランキング上位の語彙が初級語彙とは大きく異なるため、やむを得ずそうしたと説明されている(Matsushita 2012: 112–123)。4) も語彙表収録語彙の選定には語彙頻度データを活用しているが、レベルの判定は基本的にベテラン日本語教師の主観判定によるものとなっており、初級前半 424 語、初級後半 792 語、中級前半が 2,300 語、中級後半が 6,464 語、上級前半 6,379 語、上級後半 1,560 語というように、中級後半～上級前半に大量の語彙が集中し、レベル別語数の分布には著しい偏りがあり、語レベルの判定の妥当性には検討の余地がある。ほかにも「日本語共通基礎語彙」(松田ほか 2010) や「読解基本語彙 1 万語」の上位 2 千語(本田 2019) があるが、語彙表は公表されていないようである。

英語教育の世界では、長年、General Service List (通称 GSL) (West 1953) が代表的な初級語彙リストとして参照されてきた。GSL は語彙頻度などの指標を参照しながら試案が作成され、多くのフィードバックを経て、West が取りまとめたリストである。1,964 word family, 3,623 lemma からなり⁸⁾、各語につき簡単な意味・用法、及び多義語の各用法の全用法に占める割合が掲載されていて、書籍の形で出版されたものである。

しかしながら、近年、コーパス言語学の進展から、見直しの機運が出てきて、新しいリストが

⁷ 甲斐(2002: 282–283) はほかにもいくつかの文献を挙げているが、そのうちの多くは参考にするには時代が古すぎると考えられ、相対的に新しいものはその成果が「基本語二千」(国立国語研究所 1984) に反映されているため、ここでは取り上げなかった。

⁸ word family は Bauer and Nation (1993) が Level 6 として示す語の計数単位で、‘predevelopment’を‘develop’と同じ語として数えるというように、不規則かつ高頻度の接辞による派生形まで 1 語に含む単位である。lemma は、屈折変化(活用)は 1 語に含むが接辞による派生形は含まないとする単位で、develop, develops, developed, developing は同じ語と数えるが developer, development などはそれぞれ develop とは別語と数える。lemma は Bauer and Nation (1993) の定義する Level 2 に相当し、日本語語彙の計量的研究もほぼ lemma に近い単位で行われてきたと言ってよい(ただし、日本語では、接辞は語と同等の扱いを受けることが多い)。

出てきた。一つが New General Service List (通称 NGSL) (Browne et al.) で、もう一つが New-General Service List (通称 New-GSL) (Brezina and Gablasova 2015) である。

NGSL の最新版 (Version 1.2, 2023 年) は、Cambridge English Corpus から新聞及び学術のコーパスを除いた、日常生活領域の 9 つのコーパス、約 2 億 7 千万語を材料に作成されている。作成基準や方法の詳細は不明だが、コーパスの中に、文芸書や雑誌のほか、学習者コーパスや話し言葉、テレビ・ラジオ番組、ノンフィクション等が含まれていることは興味深い。このリストは GSL より多い 2,368 word family からなるが、GSL より少ない 2,818 lemma からなり、かつ GSL より高いテキストカバー率が出ると主張している。

New-GSL は British National Corpus など 4 種類の大型コーパスの計 120 億語を処理して作成されている。頻度と合わせて Average Reduced Frequency (ARF) と呼ばれる分散度指標 (ある個所に集中的に表れる語の頻度を低くする手法) を用いていることが特徴の一つである。4 種類の大規模コーパスの上位 3,000 語に共通する 2,122 語を特定して common lexical core (共通中核語彙) と呼び、さらに比較的最近よく用いられるようになった語を 2 種類の最新コーパスから選び出して、2,494 lemma を New-GSL に選定している。

これらの先行研究から示唆される重要事項は、以下のようにまとめられよう。

- ・目的に応じた、代表性の高いコーパスの使用 (特に初級に必要な日常会話コーパス)
- ・十分なコーパスサイズ (ただし、初級語彙は高頻度なので、代表性がより重要である)
- ・最新の高頻度語彙を取り入れること (大型コーパスの作成には時間がかかるため)
- ・分散度指標を用いること (初級では幅広い領域で用いられる中核的語彙が重要)
- ・選定後の、各種コーパスによるカバー率の検証、利用者のフィードバックによる改善

3.2 「基本語彙」をめぐる議論

甲斐 (2002) は「基礎語彙」「基幹語彙」等を含めた広義の「基本語彙」に関する先行研究を 1 冊の文献解題にまとめている。同書の取り上げた各種語彙リストの選定基準を検討すると、おおむね語の使用頻度、使用範囲 (幅広い領域で使われるか)、語彙の体系やまとまり (範列的 paradigmatic な関係、統合的 syntagmatic な関係にある語群、同一文脈で使われやすい語の扱い、以下「語彙体系」とする) の 3 点の扱いに集約される。

使用頻度と使用範囲についてはコーパスを扱うことで解決できるが、語彙の体系やまとまりの問題は人手による調整が必要である。範列的關係にある語については、例えば地名の場合、頻度だけを考慮すれば、ある都道府県名や国名は初級語彙で、それ以外は初級語彙ではなくなるであろう。それでよいという考え方も有力だが、別リストにしてこれらの地名を表す語彙を平等に扱うという考え方もある。また、統合的關係については、例えば「インストール」という語だけを学んで「アプリ」「ソフト」といった語を学ばないのでは「インストール」も使えないといったことが起こり得る。範列的關係でも統合的關係でもないが同一文脈で使われやすい語もある。

4.2.3 で後述する【行動・場面】【話題】の問題とつながるが、例えば、不動産屋で部屋を借りる

という状況において、「賃貸」「築〇年」「徒歩〇分」「敷金」「礼金」「契約期間」「〇〇可」「〇〇付」等の表現のうち、一部だけではうまく部屋が探せないと思われる。これらの問題を解決するには、頻度と範囲の観点から語彙を選定した後に、各語について初級用法を吟味したうえで採録語を調整し、GSLのように初級用法を提示することが望まれるであろう。

英語教育での議論を見ると、上述の諸基準のほか、West (1953) は stylistic level (フォーマル度の異なる類義語は日常語のみでよい)、intensive and emotional words (外国語学習では感情語彙はニーズが低い) について述べ、Richards (1974) は availability and familiarity (頻度と関わらず親密度の高い語も選定)⁹、coverage (類義語は一つでよく、他語で代替できない語は選定)、meaning priorities (意味そのものの必要度) も語彙選定の基準として挙げている。意味の必要度については、例えば「^{ひやくとおぼん}110番」のような危機対応語彙はそれに相当するであろうし、辞書記述に用いられるような、他の語の意味記述に最低限必要な語彙も相当するであろう¹⁰。

「基本語彙」などをめぐっては、従来は頻度や使用範囲といった語彙論的な観点からの議論が多かったが、大前提として、語彙リストそのものの利用目的、利用範囲も重要である。日常生活、学術、就労・ビジネスといった利用目的によって、学習者の必要とする語彙は初級レベルから異なるのではないだろうか。1.2 で述べたように、旧日本語能力試験出題基準で3級(現行のN4相当)とされている「研究」といった語は学術目的であれば初級かもしれないが、日常生活や就労・ビジネスでは初級から必要な語ではないであろう。このように、初級語彙の選定には、学習者の視点、すなわちニーズや場面・話題、難易度や学習順序などの視点を踏まえた検討が必要である。次節では、初級語彙の基本的条件に加えて、これらの視点も踏まえて検討する。

4. 初級語彙とはどのような語彙であるべきか

4.1 初級語彙の基本的条件

上述の諸点を踏まえると、初級語彙の特徴は以下のようにまとめられるであろう。第二言語、外国語として、様々な目的と場所でその言語を学ぶ学習者にとって、

- A) 高い頻度で出会う語彙(頻度)
- B) 幅広い領域で出会う語彙(範囲)
- C) 語の体系やまとまりとして必要な語彙(体系やまとまり)
- D) 学習が難しすぎない語彙(難易度)
- E) 言い換えの利かない語彙、すなわち他の多くの語の意味を代替的に表現するための語彙(意味的な基礎性・体系性)
- F) 危機対応に必要な語彙(他の条件に関わりなく状況によって必要度が高い語彙)
- G) 初級以降の学習の基礎となる語彙、特に機能語や基本的な内容語など

⁹ Richards (1974) が親密度について触れたのは、日常会話のコアバスや語彙頻度がほぼ参照されなかった時代のことであり、注6で述べたように、本研究では親密度は採用しないという立場をとる。

¹⁰ 語彙に用いられる語は defining vocabulary と呼ばれ、例えば Oxford Advanced Learner's Dictionary では The Oxford 3,000 と呼ばれる約 3,000 語に制限されている (Oxford Learner's Dictionaries)。

実際の設定に当たっては A, B の頻度と範囲は基本的な順位を定めるための基準であり, C ~ G は初級に入る語を加えたり削除したりする調整のための基準だと言えよう。A, B は基本的にコーパスデータを用いて決められることである。3.1 で示した New-GSL はまさにこのような考え方で語彙選定を行い, 共通中核語彙の存在を認定している。従来の初級教科書間で語彙の共通性が高くないのは語彙に十分な注意が払われなかったためかもしれないが, ニーズの高い学習項目の考慮こそが優先事項であり, それに応じてカリキュラムや教材を考えるべきである。初級教材作成の際にはシラバスの種類に関係なく, 初級で学ぶべき語彙が多く含まれているかが検討されるべきであるし, その検討に使えるような語彙リストが望まれる。

4.2 ニーズのカテゴリー化と語彙リストのモジュール化

4.2.1 言語使用の背景と学習目的

これまでに目的・対象別に行われてきた言語教育の領域を振り返ると, 日常生活のほか, 学術, 就労・ビジネス, 趣味・文化・芸術が大きな分野として存在したと考えられる。おおよそ普通の人間の生活は, 日常, 学業／仕事, 余暇からなると思われ, それらに対応する学習目的として日常生活, 学術, 就労・ビジネス, 趣味・文化・芸術の 4 領域が想定できよう。

さらに日本国内と海外では, 学習ニーズには相対的に大きな異なりがあると想定される。文化審議会国語分科会 (2019) は日本語教育人材の「活動分野」を「生活者としての外国人」「留学生」「児童生徒等」「就労者」「難民等」「海外」の 6 つに分類しているが, 「児童生徒等」は本稿では対象とせず, 「難民」になることそのものは日本語学習の目的とは異なり, 難民の学習目的は生活, および就労や留学になると考えられるので, 本稿では, 「生活者としての外国人」「留学生」「就労者」「海外」の 4 つの活動分野を扱うこととする。これらのうち, 初めの 3 つは, 上述の 4 領域のうちの日常生活, 学術, 就労・ビジネスの 3 領域と対応する。

では, 「海外」についてはどう考えたらいいであろうか。国際交流基金 (2020) が海外の日本語教育機関において 2018 年度に行った調査 (134 か国, 18,661 機関)¹¹ で, 「日本語学習の目的・理由」について 17 の選択肢を示して複数回答可で尋ねたところ, 「アニメ・マンガ・J-POP・ファッション等への興味」(66.0%), 「日本語そのものへの興味」(61.4%), 「歴史・文学・芸術等への関心」(52.4%) が上位であり, その後「日本への留学」(46.7%), 「将来の仕事・就職」(41.1%) 「日本への観光旅行」(41.1%) と続く。複数回答可なので, 上位 3 項目とそれに続く 3 項目の差に当たる約 20% の学習者は, 留学, 就職, 旅行といった実用的な目的とは異なる目的, すなわち「趣味・文化・芸術」を主な目的として日本語を学習していることになる。オーストラリアで長年日本語教育に従事したネウストプニー (2000) は日本語を生活では使わない学習者も想定すべきだと主張しており, 例えば文化・社会を知るための初級語彙といったことも想定できるであろう。海外では社会文化的な側面にウェイトが置かれる傾向が強いと言える。

ここまでの議論と国際交流基金の学習目的に関する調査の項目から, 学習目的および言語使用

¹¹ 2021 年度調査も公表されているが, コロナ禍での調査のため, 2018 年度の調査を使用した。

の背景によるニーズの相対的な違いを予想すると表1のようになる。

学習目的は細分化すれば最終的には一人一人すべて異なるということになるが、学校などで集団的に学習することを考えれば、目的や背景をある程度カテゴリー化して、その目的や背景にマッチするカリキュラムを考慮することが効率的な学習につながるであろう。

語彙の領域のカテゴリー化についても、表1のa)～d)の4種類の【学習目的】と、A)、B)の2種類の【言語使用の背景】を念頭に置いて考えればよいものと思われる。

表1 学習目的と言語使用の背景から見たニーズの相対的な違いの予想

【学習目的】	【言語使用の背景】	
	A) 日本語圏（主に日本国内）	B) 非日本語圏（主に日本国外）
a) 日常生活	多くの場面・話題に必要	将来の訪日や日本語話者の応接に必要
b) 学術	多くの専門で必要	日本語や日本関連の専門で必要
c) 就労・ビジネス	多くの職業で必要*	日本関係のビジネスや就労準備で必要
d) 趣味・文化・芸術	人により様々	日本語圏よりも相対的にニーズが高い

*職種により必要な語彙の内容と程度は非常に多様だと予想される

4.2.2 必要とされるコーパス、語彙データの種類

言語使用の背景と学習目的を合わせて考えると、初級語彙の選定に必要なコーパスは、日常会話を含む総合的なコーパス、学術系コーパス、就労・ビジネス関連のコーパス、趣味・文化・芸術（特に、アニメ・マンガ及びゲーム）に関するコーパスの4種類だと考えられる。日常生活で日本語が必要なのは基本的に日本語圏である。学術的な日本語が必要なのは日本語圏の教育機関、及び非日本語圏での日本語や日本関連の専門であり、就労・ビジネスで日本語が必要なのも多くは日本語圏であろう。非日本語圏でも日系企業等では必要であるし、就労準備でも必要だろうが、学術も就労・ビジネスも国内外を問わず、基本的に日本語圏の日本語表現を想定すればよいと思われる。趣味・文化・芸術等の比重が高くなるのは上述の通り、海外である。

コーパスを利用して初級語彙を選定するには、本来は対象者別の言語行動調査、あるいは場面・話題などのニーズ調査も必要だと思われる。すなわち、どのような媒体、場面、話題でどのような言葉をどの程度使う（と考えている）か、の調査である。すでに多くのコーパスが存在するが、コーパス自体が学習者ニーズを反映しなければ、単純に使用頻度や使用範囲に基づいてランキングを出しても役に立たない¹²。従って、言語行動調査やニーズ調査を行う場合には、その結果で語彙頻度などのデータに「重みづけ」ができるようにコーパスが対応していることが望ましい。例えば、日常生活において日常会話のニーズが高ければ、日常会話コーパスの頻度の比重を大きく考える必要がある。学術が目的であれば、学術テキストの書き言葉の比重は大きくすべきであろう。これらの点については、第5節で詳述する。

¹²「現代日本語書き言葉均衡コーパス（BCCWJ）」（約1億語）と「日本語日常会話コーパス（CEJC）」（小磯ほか2022）だけを比べても語彙は大きく異なる（Abucllil・松下2023）。従来、語彙頻度はほとんど書き言葉のコーパスしか参照されなかったことを考えると、一から見直す必要がある。CEJCの分析を見ても会話の領域により語彙の異なりは小さくない。

4.2.3 行動・場面と話題

日本で日常生活を送る場合の【行動・場面】に関しては、文化庁を中心に各種の調査が行われてきた。その結果は「『日本語教育の参照枠』に基づく『生活 Can do』一覧」（文化審議会国語分科会日本語教育小委員会 2022）という形で結実している。これは「生活上の行為の事例」について、頻度、難易度、ニーズ、すなわち「遭遇頻度」「日本語でできるか」「日本語でできるようになりたいか」に関する調査の結果（株式会社 47 ブランディング 2021）に基づいたもので、A1～B2 までの段階づけがされている。ここで「生活上の行為」とされる事項はいわゆる場面シラバスとの関係が深く、日本語教育においては、中国帰国者に対する日本語教育に始まり¹³、「リソース型生活日本語」（国際日本語普及協会）、「いろどり 生活の日本語」（国際交流基金日本語国際センター）に至る「生活者」としての日本語使用に関わる調査とその成果物に代表される。「生活 Can do」は行動・場面のニーズやレベルづけの基礎資料として有効だが、そこでどのような日本語が必要かは示されていない。しかしながら「リソース型生活日本語」「いろどり」には会話や語彙などが示されている。これらを「生活 Can do」の A1, A2 と対応させて考えることで必要な初級語彙の選定がある程度は可能だと考えられる。「リソース型生活日本語」は作成時に調査を行っており、「いろどり」も「生活 Can do」一覧作成時の場面ニーズ調査を踏まえて作成されたものなので、主観のみに基づいて作成された教科書よりも高い内容的妥当性を有するものと推測される。

ただし、そこで使用される語彙については、頻度と併せて、以下の諸点への配慮も必要である。

まず、日本語学習者の必要とする表現をコーパスだけで特定することには限界がある。日本語母語話者の使用する言語資料を主に集めた母語話者中心のコーパスは学習者ニーズを反映する保証がなく、例えばビザに関する用語など、学習者に固有のニーズもあり得る。一方で、学習者コーパスは自然な状況での大規模コーパスが存在しないいうえ、言いたいことが言えないということも想定されるため、必要な語彙を特定できないであろう。そう考えると、ニーズはコーパス以外の調査で特定したうえで、母語話者のコーパスなども活用しながら必要な語彙を特定するというプロセスが要ると思われる。その点で上述の調査に基づく「生活 Can do」と各種の語彙頻度調査を併せて見る必要がある。例えば、4-1 で触れた危機対応も「生活 Can do」に含まれており、頻度が低くてもニーズが高いことが明らかになっている。日本語圏で生活する限り初級に含まれるべきであろう。

また、【話題】は、日常会話のカテゴリーの問題であり、「おしゃべり」と一括される行動・場面に付随するが、人間関係の構築には重要で、そこで必要な語彙は話題により大きく異なる。話題が指定された大学生の会話を録音・文字化した『日本語話題別会話コーパス：J-TOCC』（中俣ほか 2021a）が公開され、これに基づいた J-TOCC 語彙表（中俣・麻 2022）も公開されている。こうした調査の活用や多様な学習者への調査の拡大も必要だと思われると同時に、語彙表を利用する現場においても、学習者の個別ニーズに基づく語彙の学習を促す必要がある。

¹³ 文化庁文化庁国語課が「中国帰国者に対する日本語教育」のサイトにその成果をまとめている。

さらに、3.2 で述べたように、【行動・場面】【話題】を扱う場合、一まとまりのやりとりに最低限必要な語彙はまとめて初級に採録するという配慮も必要であろう。

4.2.4 語彙リストのモジュール化

このように考えると、学習目的を日常会話、学術、就労・ビジネス、趣味・文化・芸術等のようにある程度カテゴリー化し、それらに対応するコーパスを、場面や話題や領域のタグをつけながら構築し、それぞれの基本語彙を独立に定めたうえで、重なりを見ていくという方法が必要だと思われる。その重なり部分が共通中核語彙、すなわち、「ハイパー初級語彙（超初級語彙）」¹⁴ となり、目的別のリストが続き、目的に応じて学習者が選択して組み合わせていくという形が考えられる。事実、Browne et al. は NGSL を提案するにあたり、語彙発達への Modular Approach を提案し、NGSL の習得の次に学習すべき語彙リストとして、Academic/Business/TOEIC/Fitness など、目的に応じて複数提案している。日本語においても学術共通語彙（松下 2011b）、文芸語彙（松下 2012）、さらには各種の異なる領域での専門語彙などが提案されているが、その前に学ぶべき肝心の初級語彙の段階から、より目的に即したモジュールの作成が必要である。そうすることで初めて全体として整合性のあるモジュールを作成できるであろう。

4.2.1 で述べたように、まず想定される語彙モジュールは、「日常生活」「学術」「就労・ビジネス」「趣味・文化・芸術」の 4 種類だと考えられる。もちろん、その先には様々な下位カテゴリーが想定される。学術であれば学術共通語彙もあるが、専門分野別モジュールも想定され、「就労・ビジネス」も職種による違いが大きいであろう。「趣味・文化・芸術」が極めて多様であることも言うまでもないが、特にアニメ・マンガやゲームはニーズが高いと思われる。これらの語彙モジュールをイメージしたのが図 1 である。

¹⁴ 目的や背景を問わないという意味で「ユニバーサル初級語彙」や中核的という意味で「コア初級語彙」といった名称も考えられる。

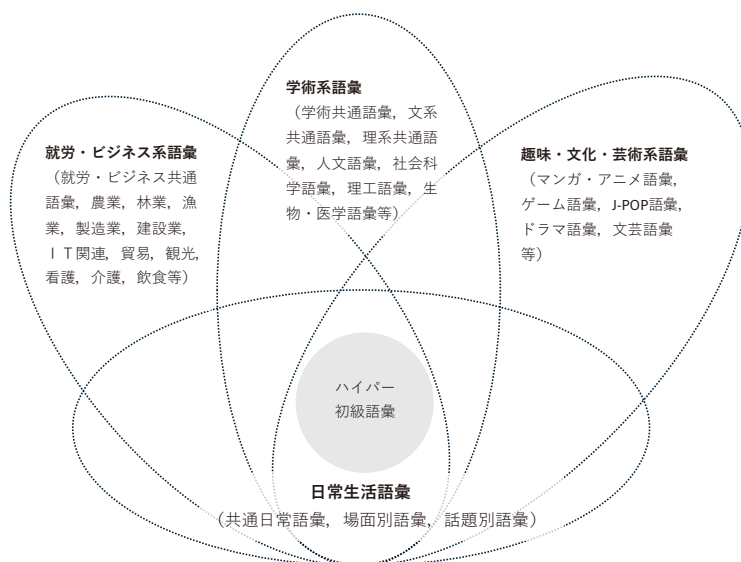


図1 語彙モジュールのイメージ

目的に応じてどのモジュールをどの順に選択していけばよいかを示せれば、学習カリキュラムを考慮するうえで有用であろう。その際、初級で学習を終えて日本語で生活する学習者の存在も想定すべきである。すなわち初級語彙の範囲である程度生活できるようにすべきだと考えられる。日本国内にも技能実習生や特定技能ビザによる一時的な滞日生活者、継続的な学習機会のない生活者など、多様なケースが存在する。加えて、上級まで進むことも想定する必要がある、中級以上へ進むための基礎作りという観点も必要であろう。

4.3 初級語彙の選定における留意点

ここまで初級語彙の条件、ニーズのカテゴリー化と語彙リストのモジュール化などについて述べてきたが、初級語彙の選定にあたっては、ほかにいくつか留意すべき点がある。

4.3.1 受容と産出の違い

「生活 Can do」(文化審議会国語分科会日本語教育小委員会 2022)において受容(聞く、読む)と産出(話す、書く)は基本的に分けられている。初級終了までのニーズを考えた場合においても、「理解できればよい語彙」と「産出できることが必要／望ましい語彙」が想定できる。しかし、語彙リスト作成に当たっては、まずは理解語彙を定めるという方針でよいと考える。理解なしで産出できることはないからである。

語彙リスト作成後には、初級レベルで書けることが望ましい漢字語を特定する必要も出てくるかもしれない。手書きが要求されるものと電子入力(スマホ／タブレットやパソコンへの入力)で済むものを分けることも考えられる。現在の社会は氏名・住所以外に手書きが要求されること

は少なく、電子入力では予測変換があるため必要なのは認識（特に同音語の書字形の識別）能力であろう。

ただし、語の難易度と漢字の難易度は一義的には別問題である。読みが難しければ仮名書きにする、ルビを振るといった対応が考えられるからである。この点は他の言語には見られない日本語に固有の問題として、別に追究する必要があるだろう。

4.3.2 難易度・学習順序の視点

使用頻度や使用範囲は社会的ニーズを反映する。高頻度で使用範囲も広い語は難度の高い語であっても重要な語であることが多い。例えば「とる」は「モノを手にとる」という用法が初級で扱われることが多いが、「壁からシール／金／BでなくA／科目／ノート／休み／バランス／場所をとる」など用法が多様で、難しい語である。

難易度で学習順序が調整される点については、国語教育の漢字の学年配当が想起される。学年配当の制定にあたり、当時の教科書での出現順や各種調査での使用頻度と使用範囲が参照されたが、他の基準も考慮され、調整を経て制定されている（丹保 2017）。例えば、「貝」が頻度順位 1,600 位以下でありながら小学校 1 年生に配当されているのは、字形の学習が比較的容易で、かつ他の漢字の部品になるからであろう。この議論からわかるのは、異なる基準を合わせて考慮すると折衷的にならざるを得ないということであり、語彙についても同様であろう。ただ、それで無原則になってよいわけではなく、まずは頻度と範囲から導出される重要度に基づいて配列し、そのうえで最低限の調整を加えることになるだろう。

4.3.3 多義語の扱い

多義語の場合、初級で身につけるべき語義と中級以上でよい語義があると想定される。派生義のほうが基本義より低頻度なら問題はないが、逆の場合はどうすべきであろうか。例えば「ハンガーに服をかける」と「電話をかける」を比べると、意味的には前者がプロトタイプの中心義だが、後者のほうが高頻度の初級表現である。「高頻度の派生義を先に教えることで問題はないか」という問いに答える用意はないが、同様の問題は、「傘をさす」の「さす」、「易しい問題」「優しい人」の「やさしい」、「うまいプレー」「うまい料理」の「うまい」など少なくない。このように、多義語の各用法の区別は、語単位のみでなく、共起成分と併せて考える必要もあろう。

語彙リストにおいて、単に語を提示するだけでなく、3.1 で述べた GSL のように、多義語の用法別頻度と用例を示せれば、この問題を解決する手掛かりにはなるであろう。

5. 異なる視点をどう統合し、語彙選定を行うか

現時点で語彙選定に使用可能と思われる語彙資料（コーパス、語彙データベース、語彙表）は表 2 の通りである。考えられる語彙の順位付け、初級語彙モジュールの選定の手順を以下に提案する。

- 1) 各コーパスの語彙頻度の性質を評価する。多次元尺度法（Multi-dimensional Scaling）によっ

て2次元上にプロットしてみることににより、各コーパスの位置づけやコーパス間の距離を直観的に把握する。あるいは特徴語抽出や高頻度語彙の観察を通して語彙表の持つ特質を把握する。

- 2) BCCWJ (書き言葉), CEJC (日常会話), jaTenTen (ウェブサイト) や、それらに近いと判定された一般的な大規模コーパスの語彙頻度を一定の割合で重みづけ (100 万語あたりの頻度にしてから決めた割合をかけて合算) して「基準頻度」を算出する。各語彙頻度の重みづけは目的とする領域 (「日常生活」「学術」「就労・ビジネス」「趣味・文化・芸術」) により変える。例えば、「日常生活」領域語彙ランクの作成にあたっては、CEJC の割合を高くし、BCCWJ の中でも Yahoo 知恵袋や文芸書の割合を高くする (方法論的妥当性は Matsushita 2012 で実証されている)。BCCWJ や CEJC のコーパス収集以降に現れたような、相対的に新しい語については、最新の語彙表である jaTenTen の頻度に大きく重みづけすることも検討する。

表2 新たな日本語初級語彙リスト作成のために参照すべき主な語彙データ

コーパス／語彙データベース／語彙表名	コーパス総語数	主な領域・用途	語彙頻度データ
a) 現代日本語書き言葉均衡コーパス (BCCWJ) 短単位語彙表 (国立国語研究所)	約 1 億語	一般	あり
b) 日本語日常会話コーパス (CEJC) 短単位語彙表 (国立国語研究所 2022)	約 240 万語	日常生活 (／ビジネス)	あり
c) jaTenTen (Sketch Engine)	約 84 億語	ウェブサイト	あり
d) 日本語を読むための語彙データベース (VDRJ) (松下 2011a)	約 3,300 万語	一般／学術／文芸	あり
e) A frequency dictionary of Japanese (FDJ) (Tono et al. 2013)	約 1 億語 + 約 750 万語	一般	あり
f) 戦後主要日本語教科書掲載語彙コーパス (田中 2018)	約 3 万語	一般?	あり
g) 日本語教育語彙表 (砂川ほか)	—	一般	なし
h) 日本語共通語彙 (松田ほか 2010)	—	一般	なし
i) 基本語二千／基本語六千 (国立国語研究所 1984)	—	一般	なし
j) 日本語教育基本語彙データベース (国立国語研究所)	—	一般	なし
k) リソース型生活日本語 (国際日本語普及協会)	—	日常生活 (場面)	なし
l) いろいろ 生活の日本語 (国際交流基金日本語国際センター)	—	日常生活 (場面)	なし
m) J-TOCC 語彙表 (中俣・麻 2022)	約 165 万語	日常生活 (話題)	あり
n) 話題別日本語語彙表 (中俣ほか 2021b)	約 113 万語	日常生活 (話題)	あり
o) (就労・ビジネス関連コーパス／語彙表)	—	就労・ビジネス	あり／なし
p) (アニメ・マンガコーパス語彙表)	—	趣味・文化	あり
q) 日本語ゲームコーパス (JGC) 語彙表 (麻 2024a, b)	約 7 千語*	趣味・文化	あり

*q) は今後の拡張により語数が変わる可能性がある

- 3) 基準頻度に分散度指標を一定の割合で掛け合わせて、領域ごとの「基準順位」を定める。分散度指標としては、VDRJ にすでに示されているものを使用するか、もしくは BCCWJ や CEJC の先コーパス頻度から改めて分散度を計算して試してみ、妥当だと思われるものを利用する。
- 4) 日本語教育用の各種の基本語彙表や「リソース型生活日本語」「いろどり」などの場面別の語彙、中俣・麻 (2022) や中俣ほか (2021b) の話題別の語彙表、松田ほか (2010) の日本語共通語彙 (親密度利用)、田中 (2016) の初級教科書の語彙頻度表などの各種の語彙表の語彙に基準頻度を当てはめてみて、これらの語彙表の語彙頻度の分布を確認する。
- 5) 「日常生活」「学術」「就労・ビジネス」「趣味・文化・芸術」の各領域において、目的に応じた語彙順位の決定後に、初級の範囲を確定するため、順位の区切り (cutoff point) を設定する。各種語彙リストの総語数、頻度の平均などの検証により区切りの設定を検討する。社会実装面からは文化庁「日本語教育の参照枠」に対応するレベル設定が望ましいかもしれない。
- 6) 「日常生活」領域の語彙については、基準頻度も参照しながら、場面タグ、話題タグなどの各種のタグをつけ、初級語彙に必要な語や不要な語を検討し、レベル付けを行う。この際、頻度だけではなく、4.1 で述べた「語の体系やまとまり」等の C ~ G の条件を考慮する。また、調整に当たっては、母語話者中心の頻度データに表れにくいものとして、ビザ関連用語のような学習者固有のニーズも考慮に入れる必要がある。
- 7) 「学術」領域の初級語彙を考える際には、VDRJ (松下 2011a) の修正版を利用して選び、それを BCCWJ の頻度と照合して確認、修正する。
- 8) 「就労・ビジネス」領域についてはデータが少ないため、今後の課題だが、「はたらくための日本語」のシリーズ (日本国際協力センター (JICE) (編著) 2019a, b), 「ゲンバの日本語」のシリーズ (一般財団法人海外産業人材育成協会 2021a, b, c, 2022a, b), 「技能実習生のための日本語『みどり』」(JITCO 日本語教材ひろば) の語彙リストなどを、許可を得て利用することも検討する。
- 9) 「趣味・文化・芸術」領域については、趣味・文化・芸術の範囲も考慮しつつ、コーパスの範囲を限定し、マンガコーパス、ゲームコーパス (麻 2024a, b) や文芸書などの語彙表も活用しながら、特に海外の日本語学習者のニーズを勘案し、「趣味・文化・芸術」領域での高頻度語を特定してタグ付けする。
- 10) 上述の過程を経てできた各種の語彙リスト試案の重なりを見ることで、語彙リストをモジュール化する。上記 4 領域の高頻度語の共通語を「ハイパー初級語彙」とするが、その後は目的に応じて「日常生活初級語彙」「学術系初級語彙」「就労・ビジネス初級語彙」「アニメ・マンガ初級語彙」などのモジュールが考えられる。
- 11) 選定後に、各種のテストコーパス (対象領域における、語彙選定時に使用していないコーパス) によるカバー率の検証や、利用者のフィードバックによる改善を行う。また、語彙項目をサンプリングして意味理解のテストを作成し、その結果からも語彙リストの妥当性を検証する。

語彙リスト作成後の普及に当たっては、社会的な影響についても責任を持たねばならない。語彙リストはあくまで共通ニーズに基づくものであり、そこに個別のニーズに応じた語彙や漢字を追加すべきことを強調する必要がある。例えば、自己紹介語彙、氏名や住所に関わる漢字、専門領域での頻出語彙などである。

6. まとめと今後の課題

研究課題 1) 「初級語彙」の選定基準については、頻度、範囲、語の体系やまとまり、難易度、など、7つの基準を 4.1 において示した。研究課題 2) 初級語彙の種類については、「日常生活」「学術」「就労・ビジネス」「趣味・文化・芸術」の 4 種類を想定し、それらのすべてに共通する「ハイパー初級語彙」も想定できるとした。研究課題 3) 「初級語彙」選定の具体的方法は、(1) コーパスの語彙頻度への目的に応じた重みづけによる順位づけ、(2) 初級の範囲の確定、(3) 体系やまとまりを考慮した調整、(4) 各種の語彙リストの重なりを見ることによる語彙のモジュール化といったステップが想定される。

これらのプロセスをまちがいになく行うことは簡単ではなく、特に就労・ビジネスや趣味・文化・芸術などは、コーパスに基づく語彙頻度データの収集自体が容易ではない。また、言語行動調査や行動・場面や話題のニーズ調査の結果を踏まえたコーパスへの重みづけの基準の策定も容易なことではないが、試案をテストコーパスのカバー率で検証することや、リストの試案へのフィードバックにより修正していくことで語彙リストを洗練されたものにしていくことはできよう。

さらには、それらの語彙リストによって、現存の教科書や教材について、語彙の側面から評価することも将来的な課題となるであろう。語彙テスト開発によるプレースメントや習得研究への応用ができるようにすることも期待したい。

様々な課題はあるが、社会の変化に対応した新たな初級語彙リストのモジュールの構築は日本語教育界における喫緊の課題の一つである。本稿がその一助となれば幸いである。

参考文献

- Abuella, Nora・松下達彦 (2023) 「日本語の日常会話の「形式×場所」「性別×年齢」別の語彙の特徴—日本語日常会話コーパス (CEJC) の多次元尺度法と特徴語抽出による分析—」国立国語研究所共同研究発表会 (2023 年 1 月 8 日) 発表資料。
- Bauer, Laurie and Paul Nation (1993) Word families. *International Journal of Lexicography* 6: 253–279.
- Brezina, Vaclav and Dana Gablasova (2015) Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics* 36: 1–22.
- Matsushita, Tatsuhiko (2012) In what order should learners learn Japanese vocabulary? A corpus-based approach. Doctoral dissertation, Victoria University of Wellington. https://openaccess.wgtn.ac.nz/articles/thesis/In_What_Order_Should_Learners_Learn_Japanese_Vocabulary_A_Corpus-based_Approach/17011514?file=31466030 (2024 年 12 月 9 日確認)
- Matsushita, Tatsuhiko (2023) Text covering efficiency and word tier analysis for the proposal of vocabulary learning order and the analysis of text genres. Poster presented at QUALICO 2023. http://www17408ui.sakura.ne.jp/tatum/presentation/Matsushita2023b_poster_TCE_medical.pdf (2024 年 12 月 9 日確認)
- Richards, Jack C. (1974) Word lists: Problems and prospects. *RELC Journal* 5(2): 69–84.
- Tono, Yukio, Makoto Yamazaki and Kikuo Maekawa (2013) *A frequency list of Japanese*. London, New York: Routledge.

- West, Michael (1953) *A general service list of English words*. London: Longman.
- 天野成昭・近藤公久 (2000) 『日本語の語彙特性』第1巻. 東京: 三省堂.
- 一般財団法人海外産業人材育成協会 (2021a) 『ゲンバの日本語 基礎編 働く外国人のための日本語コミュニケーション』東京: スリーエーネットワーク.
- 一般財団法人海外産業人材育成協会 (2021b) 『ゲンバの日本語 応用編 働く外国人のための日本語コミュニケーション』東京: スリーエーネットワーク.
- 一般財団法人海外産業人材育成協会 (2021c) 『ゲンバの日本語 単語帳 製造業 働く外国人のためのことば』東京: スリーエーネットワーク.
- 一般財団法人海外産業人材育成協会 (2022a) 『ゲンバの日本語 単語帳 IT 働く外国人のためのことば』東京: スリーエーネットワーク.
- 一般財団法人海外産業人材育成協会 (2022b) 『ゲンバの日本語 単語帳 建設・設備 働く外国人のためのことば』東京: スリーエーネットワーク.
- 伊東祐郎 (2006) 「評価の観点から見た日本語教育スタンダード」『日本語学』25: 18-25.
- 今西利之・神崎道太郎 (2008) 「日本語教育初級教科書提示語彙の数量的考察」『熊本大学留学生センター紀要』4: 1-20.
- 岩田一成・森篤嗣・松下達彦 (2015a) 「『やさちチェック』—公的文書の難易度を指標化するシステムの開発—」『2015年度日本語教育学会秋季大会 予稿集』407-408.
- 甲斐睦朗 (2002) 「現代日本語の基本語彙」飛田良文・佐藤武義 (編) 『現代日本語講座 語彙』第4巻. 25-45. 東京: 明治書院.
- 株式会社 47 ブランディング (2021) 「文化庁委託事業報告書 令和2年度 日本語教育総合調査補助業務「生活者としての外国人」に対する日本語教育の標準的なカリキュラム案」の改定のための基礎調査及び Can do 作成に係る事業報告書」https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/nihongokyoiku_sogo/r02/pdf/93685801_03.pdf (2024年12月9日確認)
- 北條淳子 (1989) 「中・上級の指導上の問題」寺村秀夫 (編) 『講座日本語と日本語教育』13: 238-267. 東京: 明治書院.
- 木村宗男・阪田雪子・窪田富男・川本喬 (1989) 『日本語教授法』東京: 桜楓社.
- 小磯花絵・天谷晴香・石本祐一・居關友里子・白田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香 (2022) 『『日本語日常会話コーパス』設計・構築・特徴』(国語研究所「日常会話コーパス」プロジェクト報告書 6).
- 国際交流基金 (2020) 『海外の日本語教育の現状 2018年度日本語教育機関調査より』<https://www.jpff.go.jp/j/project/japanese/survey/result/survey18.html> (2024年12月9日確認)
- 国際交流基金・日本国際教育協会 (1994) 『日本語能力試験出題基準』東京: 凡人社.
- 国際文化振興会 (1944) 『日本語基本語彙』東京: 国際文化振興会.
- 国立国語研究所 (1962) 『現代雑誌九十種の用語用字第1分冊: 総記および語彙表』東京: 秀英出版.
- 国立国語研究所 (1984) 『日本語教育のための基本語彙調査』東京: 秀英出版.
- 佐藤浩史・笠原要・金杉友子・天野成昭 (2004) 「単語親密度に基づく基本語彙の選定」『人工知能学会論文誌』19: 502-510.
- 獅々見真由香 (2016) 「日本語の会話におけるオノマトペの基本語彙選定」『日本語教育』165: 73-88.
- 田中祐輔 (2016) 「初級総合教科書から見た語彙シラバス」森篤嗣 (編) (2016) 第1章. 3-31.
- 丹保健一 (2017) 「学年別漢字配当表の字種選定に関する基礎的研究—戦後の漢字表作成に見られる字種選定基準の諸相」『三重大学教育学部研究紀要』68: 47-60.
- 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣 (2021a) 「『日本語話題別会話コーパス: J-TOCC』」『計量国語学』33(1): 11-21.
- 中俣尚己・小口悠紀子・小西円・建石始・堀内仁 (2021b) 「自然会話コーパスを基にした『話題別日本語語彙表』」『計量国語学』33(3): 194-204.
- 中俣尚己・麻子軒 (2022) 「『日本語話題別会話コーパス: J-TOCC 語彙表』の公開と日本語教育むけ情報サイトにむけた指標の検討」『言語資源ワークショップ 2022 発表論文集』https://clrd.ninjal.ac.jp/lrw/lrw2022/p1-4_paper.pdf
- 日本国際協力センター (JICE) (編著) (2019a) 『はたらくための日本語 職場の語彙と表現』I/II/III. 東京: 凡人社.
- 日本国際協力センター (JICE) (編著) (2019b) 『はたらくための日本語 職場のコミュニケーション』I/II/III. 東京: 凡人社.

- ネウストブニー, J.V. (2000) 『今日と明日の日本語教育—21 世紀のあけぼのに』 東京: アルク.
- 本田ゆかり (2019) 「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: 118-133.
- 麻子軒 (2024a) 『『日本語ゲームコーパス (JGC)』の構築に関する中間報告: 前期のアクションゲームに見られる量的特徴』『言語資源ワークショップ発表論文集』279-287. <https://doi.org/10.15084/0002000371>
- 松下達彦 (2016) 「コーパス出現頻度から見た語彙シラバス」森篤嗣 (編) (2016) 第 3 章, 53-77.
- 松田真希子・児玉茂昭・竹元勇太・石坂達也・森篤嗣・川村よし子・山本和英 (2010) 「コーパスの異なりと単語親密度を活用した日本語共通語彙の抽出」『言語処理学会第 16 回年次大会発表論文集』579-582.
- 森篤嗣 (編) (2016) 『ニーズを踏まえた語彙シラバス』東京: くろしお出版.
- 森田良行 (1986) 「初-中級移行過程における語彙教育」『講座日本語教育』22: 98-108. 早稲田大学語学教育研究所.
- 吉島茂・大橋理枝 (訳・編) (2004) 『外国語教育 II—外国語の学習, 教授, 評価のためのヨーロッパ参照枠—』東京: 朝日出版社.

関連 Web サイト (いずれも 2024 年 12 月 9 日確認)

- Browne, Charles, Brent Culligan and Joseph Phillips. *New General Service List Project*. <https://www.newgeneralservicelist.com/>
- Cambridge University Press & Assessment (2024) *English Vocabulary Profile Online*. <https://englishprofile.org/wordlists/evp>
- JITCO 日本語教材ひろば「技能実習生のための日本語「みどり」」<https://hiroba.jitco.or.jp/categories/index/1>
- Lee, Jae-ho and Yoichiro Hasebe (2013) 「jReadability 日本語文章難易度判別システム」<https://jreadability.net/sys/ja>
- Oxford Learner's Dictionaries. *The Oxford 3000*. https://www.oxfordlearnersdictionaries.com/wordlist/american_english/oxford3000/
- Sketch Engine. jaTenTen: Corpus of the Japanese Web. <https://www.sketchengine.eu/jatenten-japanese-corpus/>
- 岩田一成・森篤嗣・松下達彦・中島明則 (2015b) 「やさになちチェッカー」<http://www4414uj.sakura.ne.jp/Yasanichi1/nsindan/>
- 川村よし子・北村達也・保原麗「日本語読解学習支援システム リーディング チュウ太」<https://chuta.cegloc.tsukuba.ac.jp/>
- 国際交流基金「JF 日本語教育スタンダード」<https://www.jfstandard.jp/f.go.jp/top/ja/render.do>
- 国際交流基金・日本国際教育支援協会「旧試験との比較」<https://www.jlpt.jp/about/comparison.html>
- 国際交流基金日本語国際センター「いろいろ 生活の日本語」<https://www.irodori.jpf.go.jp/index.html>
- 国際日本語普及協会「リソース型生活日本語」<https://www.ajalt.org/resource/>
- 国立国語研究所「現代日本語書き言葉均衡コーパス 短単位語彙表 (Version 1.1)」<https://doi.org/10.15084/00003219>
- 国立国語研究所「日本語教育基本語彙データベース」<https://mmsrv.ninjal.ac.jp/brfvep/>
- 国立国語研究所 (2022) 「日本語日常会話コーパス CEJC 短単位語彙表 _ 語彙素のみ _ 語形別 ver 202209」<https://www2.ninjal.ac.jp/conversation/cejc/cejc-wc.html>
- 菅長陽一・松下達彦 (2013) 「日本語テキスト語彙分析器 J-LEX」<http://www17408ui.sakura.ne.jp/index.html>
- 砂川有里子・李在鎬・川村よし子・今井新悟・杉本武・長谷部陽一郎・高原真理「日本語教育語彙表」<https://jhlee.sakura.ne.jp/JEV/>
- 田中祐輔 (2018) 「戦後主要日本語教科書掲載語彙コーパス」森篤嗣 (編) 『コーパスで学ぶ日本語 日本語教育への応用』東京: 朝倉書店, 第 2 章 web 資料. https://www.asakura.co.jp/detail.php?book_code=51655&srsltid=AfmBOOpYHeLXGePczlK8uSizy3-MMqvufAT_cyCuWZlkEOVV9h6NWXcW
- 文化審議会国語分科会 (2019) 「日本語教育人材の養成・研修のあり方について (報告) 改訂版」https://www.bunka.go.jp/seisaku/bunkashingikai/kokugo/kokugo/kokugo_70/pdf/r1414272_04.pdf
- 文化審議会国語分科会 (2021) 「日本語教育の参照枠」https://www.bunka.go.jp/seisaku/bunkashingikai/kokugo/hokoku/pdf/93476801_01.pdf
- 文化審議会国語分科会日本語教育小委員会 (2022) 「地域における日本語教育の在り方について (報告) 別冊『日本語教育の参照枠』に基づく『生活 Can do』一覧」https://www.bunka.go.jp/seisaku/bunkashingikai/kokugo/hokoku/pdf/93913301_01.pdf

- 文化庁文化庁国語課「中国帰国者に対する日本語教育」https://www.bunka.go.jp/seisaku/kokugo_nihongo/kyoiku/china_kikokusha/
- 本田ゆかり「読解基本語彙チェッカー」<https://basic.chuta.jp/>
- 麻子軒 (2024b)「『日本語ゲームコーパス (JGC)』 語彙表 (前期アクションゲーム)」<https://kenjima.net/download.html>
- 松下達彦 (2011a)「日本語を読むための語彙データベース (VDRJ) Ver. 1.0」<http://www17408ui.sakura.ne.jp/tatsum/database.html#vdrj>
- 松下達彦 (2011b)「日本語学術共通語彙リスト Ver.1.01」<http://www17408ui.sakura.ne.jp/tatsum/list.html#jcaw>
- 松下達彦 (2012)「日本語文芸語彙リスト Ver. 1.00」<http://www17408ui.sakura.ne.jp/tatsum/list.html#jlw>

How Should Elementary Vocabulary be Defined in Teaching Japanese as a Second or Foreign Language?

MATSUSHITA Tatsuhiko^a IWASHITA Tomohiko^b ARAI Tomohiro^c
 ABUELLIL Nora^d TANAKA Yusuke^e NAKAMATA Naoki^f
 SRDANOVIĆ Irena^g MATSUDA Makiko^h LIU Ruiliⁱ CHEN Mengxia^j
 KASHINO Wakako^a ISHIGURO Kei^a HASHIMOTO Naoyuki^k OGUMA Rie^l

^aNINJAL / The Graduate University for Advanced Studies, SOKENDAI

^bGraduate Student, University of Electro-Communications /

Adjunct Researcher/ Project Collaborator, NINJAL

^cGraduate Student, Meiji University / Adjunct Researcher/ Project Collaborator, NINJAL

^dGraduate Student, Waseda University / Project Collaborator, NINJAL

^eUniversity of Tsukuba / Project Collaborator, NINJAL

^fThe University of Osaka / Project Collaborator, NINJAL

^gJuraj Dobrila University of Pula / Project Collaborator, NINJAL

^hTokyo Metropolitan University / Project Collaborator, NINJAL

ⁱSun Yat-sen University / Project Collaborator, NINJAL

^jNingbo University / Project Collaborator, NINJAL

^kFukuoka Women's University / Project Collaborator, NINJAL

^lCatholic University of Louvain / Project Collaborator, NINJAL

Abstract

Presenting elementary vocabulary in the form of a list is a significant exercise; these lists are then used to create teaching materials and vocabulary tests. However, the vocabulary lists from Levels 4 and 3 of the former Japanese Language Proficiency Test, which are among the most commonly referenced, no longer suit to the changes in society, necessitating a reassessment of elementary vocabulary lists. This study aims to explore the ideal structure of such vocabulary lists. It begins by examining the issues with existing lists and reviewing discussions and outcomes related to basic vocabulary selection, including studies on English vocabulary lists. Subsequently, it identifies seven key conditions for elementary vocabulary, namely frequency, range, and lexical system, and advocates the creation of domain-specific vocabulary modules tailored to 1) everyday life, 2) academics, 3) work and business, and 4) hobbies, culture, and arts. Moreover, the study also identifies vocabulary resources that can be used to select elementary vocabulary and outlines an 11-step procedure for vocabulary selection and validation. These steps include evaluating the characteristics of lexical frequency in corpora, applying weighting to frequency data, determining standard rankings while considering dispersion, and confirming the frequency and range of vocabulary within lists. Furthermore, this study highlights the key considerations in creating vocabulary lists, such as addressing the needs of overseas learners, incorporating perspectives on situational and topical contexts, distinguishing between importance and difficulty, differentiating receptive and productive vocabulary, and managing issues related to polysemous words, thereby providing a structured framework for developing these lists.

Keywords: elementary vocabulary list, frequency, range, lexical system, vocabulary module