

日本語教育での語彙資源の活用 ―現状と展望―

松下達彦(国立国語研究所)

* フォーラム終了後に一部加筆 (2023 年 2 月 20 日)

1. 日本語教育で活用されてきた語彙資源(歴史・現状)

学習順序、学習レベルを考えるうえで、頻度データが重要。頻度は社会的ニーズ(=重要度)を示す。難易度とは違うことに注意が必要。

1.1 戦前～1984

- ・基本語彙表に近いものは大正時代から存在(国立国語研究所, 2000)
- ・『現代雑誌九十種の用字・用語』(語彙表)(国立国語研究所, 1962)の時代
唯一の大規模頻度データとして参照された

1.2 1984～1994

- ・『日本語教育基本語彙七種比較対照表』(国立国語研究所, 1982)から『日本語教育のための基本語彙調査』(国立国語研究所, 1984)へ

1.3 1994～2010 頃

- ・『日本語能力試験出題基準』(1994, 2002): 初級(4 級、3 級)は教科書 11 種比較対照データを利用、短単位でない句レベルの表現も。中上級(2 級、1 級)は国立国語研究所(1984)も参照。
⇒語彙資料に基づいて教科書ができたのではなく、教科書が語彙資料として参照されていた。初級については、語彙の計数に使えるよい会話コーパスがなかったことが原因。
⇒対策本と「リーディング チュウ太」(Kawamura et al., 1997)の多大な影響
- ・パソコンの普及と研究の増大: 天野・近藤(2000)、国立国語研究所(2006)等。教材や教授法への影響は不明(日本語教育研究には活用されたが、教育そのものへの応用は、直接的にはあまりないかも・・・)
- ・『児童生徒に対する日本語教育のための基本語彙調査』(工藤, 1999)は国立国語研究所(1984)「基本語 2000」も参照。
- ・『品詞別・レベル別 1 万語語彙分類集』(専門教育出版テスト課編, 2001)
- ・紙の辞書の衰退、スマホアプリの多大な影響: WWWJDIC (EDRDG, 1997～)が多くの辞書に

1.4 2010 頃～現在 (2023)

- ・2010 年、日本語能力試験の新試験の登場、出題範囲の語彙表は改訂され、非公開に: 範囲は 15,000 語以上に拡大(秋元・押尾, 2008)
- ・UniDic(伝ほか, 2007～), 「現代日本語書き言葉均衡コーパス(BCCWJ)」 「教科書コーパス」(国立国語研究所, 2011)の登場
⇒各種の語彙表やツールへの応用: 松下(2011a, b), 本田(2015, 2019a, b)等
- ・「日本語教育語彙表」(2015): 約 18,000 語。複数日本語教師の主観判定(Sunakawa et al., 2012)
- ・「J-TOCC 語彙表」(中俣・麻, 2022): 「日本語話題別会話コーパス(J-TOCC)」(中俣, 2021; 中俣ほか, 2021)利用
- ・「日本語日常会話コーパス(CEJC)」(国立国語研究所, 2022)の登場

2. 語彙資源はどう活用されているか(応用の現状) *松下(2017)を参照

- A) 語彙シラバスに基づく教材やコースの開発: 簡単ではない。試験対策など自習用が多い。
石澤ほか(2018, 2021)『語彙ドン』1, 2 ←「日本語学術共通語彙リスト」(松下, 2011b)

- B) 語彙データベース、特徴語リスト：チェックリストとしての活用、教材・語彙学習ツール（スマホアプリなど）への活用
 ⇒日本語能力試験出題範囲（現行試験：2010～、語彙表非公開）
 「小中学生のための日本語学習語リスト」（バトラー後藤，2010）：「教科書コーパス」（最終成果：国立国語研究所，2011）等を利用
 「日本語を読むための語彙データベース」（松下，2011a）
 「日本語学術共通語彙リスト」（松下，2011b）、「日本語文芸語彙リスト」（松下，2012）
 （「留学生のための基礎的専門語データ」（小宮，2014，2018）は専門書データ利用）
- C) 語彙頻度プロファイラー等 ＊裏に形態素解析辞書やデータベースが入っている
 ＝文章の語彙的負荷の測定。既知語のテキストカバー率と文章の読解レベルの関係の研究がベース。
 学習負荷・学習効率を調節するリライトへの応用。
 「リーディングチュー太」（Kawamura, et al., 1997）
 「やさしにちチェッカー」（岩田ほか，2012）：語彙部分は松下（2011a）を利用
 「J-LEX」（Suganaga & Matsushita, 2013）：語彙部分は松下（2011a）を利用
 「jReadability 日本語文章難易度判別システム」（Lee & Hasebe, 2013）
 「読解基本語彙チェッカー」（本田，2019b）
 「やんしす」（YAsashii Nihongo SIen System）（伊藤，2008）
- D) コロケーション（連語）・例文の検索
 「NINJAL-LWP for BCCWJ (NLB)」（国立国語研究所・Lago 言語研究所，2012）：約 1 億語
 「NINJAL-LWP for TWC (NLT)」（筑波大学・国立国語研究所・Lago 言語研究所，2013）：約 11 億語
 ⇒2 語比較あり・・・例文の質は NLB が上だが、低頻度語は NLT のほうが信頼度が高い
 「なつめ」：理工系学会誌論文や Wikipedia をコーパスとして利用
- E) 語彙テスト ＊ウェブで無料受験できる診断用のテスト
 「日本語学術共通語彙テスト（音声版／文字版）」（佐藤ほか，2017）：松下（2011a）利用
 「日本語あなうめテスト JACKET (Japanese Cloze Key Test)」：松下（2011a）利用
 「JaVocab」（篠崎ほか，2018）：松下（2011a）利用
 「語彙数推定テスト(令和版)」（NTT コミュニケーション科学基礎研究所，n.d.）：単語親密度データベースに基づく
 「ごいす〜」（読解力研究会，2021）：松下（2011a）利用
- F) リーダビリティ測定 ＊裏に形態素解析辞書やデータベースが入っている
 「帯 3」（名古屋大学佐藤理史研究室，2008）（現在使用不可か）
 「日本語リーダビリティ測定」（Shibasaki, 2017）

3. どのような語彙資源が期待されるか、日本語教育にどう活用されるか(期待と展望)

- 3.1 とりあえず＝現下の期待 （個人的な見解で、国立国語研究所としての見解ではありません）
- ・「日本語日常会話コーパス(CEJC)」語彙表の応用：例えば、NINJAL-LWP への搭載
 - ・学習者の背景・目的の違いに基づいた語彙データの作成（データベースへのタグ付け）
 特に初級語彙リストの見直し（現在、進行中）
 - ・さらなる大規模な語彙データ → 低頻度語彙の語彙資源の高精度化
 - ・長単位リストの開発と応用，長単位データと短単位データの融合？
 - ・文章や会話の語彙的難度の簡便な測定ツールの開発
 - ・語彙テスト開発：多様な側面の測定 例) Vocabulary Levels Test (Webb et al., 2017), Word Part Levels Tests (Sasao & Webb, 2017) の日本語版、音声版

- ・ 日日翻訳ツールの開発：「やさしい日本語」作成の自動化
 ……語彙資源というよりも書き換えの学習データが必要か、誰がどう使うか
- ・ 新語リスト：属性情報（レジスター情報：若者語、ネット用語、等）
- ・ 方言/地域語データの活用。共通語との対応語彙リストがあれば活用しやすいか。
- ・ 学齢期前の基本語彙リスト：移民の子どものサポート

3.2 テクノロジーの進展と語彙資源の応用＝近未来の展望

- ・ ER, VR の応用 例）ポケモン Go のような語彙学習ゲーム、[INFLUENT](#)
- ・ 文章→対訳付き語彙リスト＋練習問題の自動作成（AI の応用？）…学習者用、教師用
- ・ 翻訳機の発展形？

*語彙資源の活用には皆が使えるものを作る必要がある。わかりやすいのは教材、学習ツール等への応用。

*ツールが高度化すれば言語教育の対象・目的が変化。その時に必要な語彙資源とは？

<参考文献>

- 秋元美晴・押尾和美（2008）「新しい日本語能力試験のための語彙表・漢字表作成中間報告 一新語彙表 Ver. III の完成まで」『日本語学』27(10), 36-49.
- 天野成昭・近藤公久（2000）『日本語の語彙特性 第2期』三省堂
- 石澤徹・岩下真澄・伊志嶺安博・桜木ともみ・松下達彦（2018）『語彙ドン！ 大学で学ぶためのことば』vol.1、くろしお出版
- 石澤徹・岩下真澄・桜木ともみ・松下達彦（2021）『語彙ドン！ 大学で学ぶためのことば』vol.2、くろしお出版
- 伊藤彰則（2008）[「やんしす」\(YAsashii Nihongo SIen System\)](#)
- 岩田一成・森篤嗣・松下達彦・中島明則（2012）[『やさしにちチェッカー』](#)
- 工藤真由美（1999）『児童生徒に対する日本語教育のための基本語彙調査』ひつじ書房
- 佐藤尚子・田島ますみ・松下達彦・笹尾洋介・橋本美香（2017）[「日本語学術共通語彙テスト（音声版／文字版）」](#)
- 国際交流基金・日本国際教育協会編（1994）『日本語能力試験出題基準』凡人社
- 国際交流基金・日本国際教育協会編（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（1962）[『現代雑誌九十種の用字・用語 第一分冊 総記および語彙表』](#) 秀英出版
- 国立国語研究所（1982）[『日本語教育基本語彙七種比較対照表』](#) 大蔵省印刷局
- 国立国語研究所（1984）[『日本語教育のための基本語彙調査』](#) 秀英出版
- 国立国語研究所（2000）[『日本語基本語彙 一文献解題と研究一』](#) 明治書院
- 国立国語研究所（2006）[『現代雑誌 200 万字言語調査語彙表』](#)
- 国立国語研究所（2011）[「特定領域研究『日本語コーパス』言語政策班最終成果 CD-ROM」](#)
- 国立国語研究所（2022）[「日本語日常会話コーパス（CEJC）」](#)
- 国立国語研究所・Lago 言語研究所（2012, 2016）[NINJAL-LWP for BCCWJ \(NLB\)](#)
- 小宮千鶴子（2014）[「留学生のための経済の基礎的専門語」](#)『早稲田日本語研究』23, 1-12.
- 小宮千鶴子（2018）[「留学生のための物理の基礎的専門語」](#)『早稲田日本語研究』27, 37-48.
- 小宮千鶴子（2018）[「留学生のための化学の基礎的専門語」](#)『専門日本語教育研究』20, 43-48.
- 篠崎大司・川村よし子・松下達彦（2018）[「JaVocab」](#)
- 専門教育出版テスト課編（2001）『品詞別・レベル別 1 万語語彙分類集』凡人社
- 筑波大学・国立国語研究所・Lago 言語研究所（2013, 2021）[NINJAL-LWP for TWC \(NLT\)](#)
- 伝康晴・山田篤・小椋秀樹・小磯花絵・小木曾智信（2007, 2022）[UniDic.](#)
- 読解力研究会（2021）[「ごいす〜」](#)
- 中俣尚己（2021）[「日本語話題別会話コーパス：J-TOCC 解説資料」](#)
- 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣（2021）「『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・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向け情報サイトにおけ](#)

[「指標の検討」](#)『言語資源ワークショップ2022 発表論文集』

日本語教育語彙表 (2015) <http://jhlee.sakura.ne.jp/JEV/>

バトラー後藤裕子 (2010) [「小中学生のための日本語学習語リスト \(試案\)」](#)『母語・継承語・バイリンガル教育 (MHB) 研究』6, 42-58.

本田ゆかり (2015) [「大規模コーパスに基づく日本語教育語彙表の作成」](#) 博士論文 (東京外国語大学)

本田ゆかり (2019a) [「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.

本田ゆかり (2019b) [「読解基本語彙チェッカー」](#)

松下達彦 (2011a) [「日本語を読むための語彙データベース」](#) Ver. 1.01.

松下達彦 (2011b) [『日本語学術共通語彙リスト』](#)

松下達彦 (2012) [『日本語文芸語彙リスト』](#)

松下達彦 (2017) 「語彙リストの利用法ーコーパス分析に基づく語彙研究は何を目指すべきかー」『専門日本語教育研究』19, pp.19-24.

松下達彦 (2021) [「日本語あなうめテスト JACKET \(Japanese Cloze Key Test\)」](#)

EDRDG (Electronic Dictionary Research and Development Group)(2021). (1997, 2022). [Jim Breen's WWWJDIC](#).

Kawamura, Y., Kitamura, T. & Hobara, R. (1997). [「リーディング チュウ太」](#)

Lee, J. & Hasebe, Y. (2013). [「jReadability 日本語文章難易度判別システム」](#)

Sasao, Y. & Webb, S. (2017). The Word Part Levels Test. *Language Teaching Research*, 21(1), 12-30.

Shibasaki, H. (2017). [「日本語リーダビリティ測定」](#)

Suganaga, Y. & Matsushita, T. (2013) [J-LEX](#)

Sunakawa, Y., Lee, J., & Takahara, M. (2012). [The Construction of a Database to Support the Compilation of Japanese Learners Dictionaries](#), *Acta Linguistica Asiatica* 2(2), 97-115.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics*, 168(1), 34-70.