

国立国語研究所学術情報リポジトリ

「日本語非母語話者の読解コーパス」から見える非漢字圏日本語学習者の辞書使用

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2020-02-06 キーワード: 日本語非母語話者の読解コーパス 作成者: フメリヤク寒川, クリスティナ, Hmeljak, Sangawa Kristina メールアドレス: 所属:
URL	https://doi.org/10.15084/00002587

「日本語非母語話者の読解コーパス」から見える 非漢字圏日本語学習者の辞書使用

クリスティナ・フメリャク寒川（リュブリャナ大学／国立国語研究所外来研究員）[†]

Dictionary Use by Non-Kanji Background Learners of Japanese as Seen in the “Reading Corpus of Non-Native Speakers of Japanese”

Kristina Hmeljak Sangawa (University of Ljubljana / NINJAL visiting researcher)

要旨

非漢字圏日本語学習者が日本語を読む際、読めない漢字や分からない言葉を辞書で調べるのに相当な時間を費やすが、学習者がどんな辞書をどのように使い、そこでどんな工夫と困難点がみられるかを探るために、国立国語研究所で開発中の「日本語非母語話者の読解コーパス」に収録されたデータの一部を、辞書使用という観点から分析した。このコーパスのデータは、日本語学習者が普段使っている辞書などのツールを使いながら各自が選んだ文章を読み、理解したことを母語で話す場面を録音（一部録画）、文字化したデータである。調査対象はヨーロッパ各地の大学で日本語を習っている初級から上級までの学習者29名のデータである。分析の結果、辞書選びの段階から、検索を実行し、得られた情報を文脈へ適用する段階まで様々な工夫と困難点がみられたが、より効果的な辞書使用の指導に向けた辞書使用分析の枠組みとデータアノテーションについて述べる。

1. はじめに

日本語学習者が日本語の文章を読むとき、学習者用に語彙が制限された文章を除けば、文章を理解するために辞書を頻繁に使用する。特に非漢字圏の学習者は、読めない漢字や分からない言葉を探すのに多くの時間と労力を費やす。しかし、辞書の選び方についても、その使い方についても、日本語の授業で紹介されることは少ない。多くの場合、日本語学習者は辞書についての授業を受けずに自ら辞書を選び、そして試行錯誤を重ねてその使い方を覚えていく。その結果、自分のニーズに合う辞書を手に入れ、使いこなせるようになる学習者もいるが、それができず、漢字や語彙を調べるのに苦労しつづける学習者もいる。

そこで、このような学習者の辞書使用を指導するために、非漢字系日本語学習者が日本語を読むとき、どんな工夫が文章の理解に役立つか、どの困難点はその理解を妨げるかを明らかにする必要がある。この調査では、非漢字系日本語学習者の辞書使用の実態を探るために、国立国語研究所で開発中の『日本語非母語話者の読解コーパス』の初級から上級までの学習者の辞書使用データを分析した。その結果、辞書を選ぶ段階から、検索法を決めて言葉や漢字を探す段階、そして辞書から得られた情報を文脈に適用する段階まで、辞書使用過程のそれぞれの段階において効果的な工夫も見られたが、困難な場面も少なくなかった。以下の節では、調査データと調査方法について述べ、それぞれの段階の主な困難点を紹介し、最後に、調査の限界と今後の課題について述べる。

[†] kristina.hmeljak@ff.uni-lj.si

2. 調査のデータ

この調査に使ったデータは、国立国語研究所で開発中の『日本語非母語話者の読解コーパス』(野田 2019、以下『読解コーパス』と略す)に収録されたデータの一部である。『読解コーパス』の調査では、日本語非母語話者に日本語の文章を普段のように辞書などを使いながら読んでもらい、理解したことを母語(または母語に準ずる言語)で話してもらい、その発話を録音(一部のデータは録画)している。学習者の読解に立ち会うデータ収集者は、学習者の発話を聞き、そして、その理解過程を確認する必要があると判断した場合、学習者に質問し、答えてもらう。読解の途中、学習者が辞書などのリソースを使ったとき、学習者がどの文字を見て、どのような検索過程を経て、どのような意味を理解したかを、データ収集者が記録する。最後に録音したデータを文字化し、学習者の発話を日本語に訳し、そのデータを公開する。

このように収集したデータの内、『読解コーパス』のサイトにおいて公開されているのは、a) 学習者が読んだ文章、b) 学習者とデータ収集者の発話の文字化と日本語訳、c) 学習者についてのメタデータ(日本語学習歴、日本語力のレベル、使用した辞書など)である。録音ファイル及び録画ファイルは、未公開のサイトにおいてデータ収集者のみが共有する。このファイルに学習者が普段どの辞書を使い、なぜその辞書を選んだかなどの情報が一部のデータに記録されている。

本調査では『読解コーパス』のサイトにおいて公開されているデータの他に、辞書使用を詳細に観察するために、データ収集関係者が共有する未公開のサイトのデータも利用した。使用したデータは「読解コーパス」の調査方法に基づいて2016年12月から2019年1月にかけて得られたデータである。調査対象となった学習者はヨーロッパ各地の大学で日本語を習っている初級から上級までの学習者29名である。その母語はスロベニア語14名(内1名はマケドニア語とのバイリンガル)、スペイン語6名(内3名はカタルーニャ語とのバイリンガル)、フランス語6名(内1名は英語とのバイリンガル)、ドイツ語2名、クロアチア語1名である。

3. 調査枠組み：辞書使用過程の段階

読めない漢字、わからない単語を辞書で調べることは一見単純な作業だが、必要な情報を得てそれを文章の理解に役立てるためには、表1に示すように、いくつかの段階を踏まなければならない。この表はHartmann(1989)の辞書検索フローチャートを、電子辞書やオンラインツールなども含めた日本語読解過程の辞書使用に合わせて改変したものである。学習者はわからない言葉に遭遇した度に、この段階を順番に踏み、このプロセスを繰り返していく。

表1 辞書使用過程の流れ

検索段階	検索行為
1. 読解環境の整備	辞書・リソース・ツールを選択、入手
2. 検索対象の特定	a. 文章を区切り検索項目を特定 b. 検索すべきかどうかを判断
3. 検索方法の決定	a. 辞書形(検索する見出し)を決定 b. 辞書、検索方法を選択
4. 検索実行	a. 見出しの検索を実行 b. 辞書項目から必要な情報を抽出
5. 検索結果の適用	検索結果を文脈に適用

2で述べたデータを分析するために、表1に示した段階に沿って、学習者が読解において辞書を使うプロセスをたどり、その工夫と困難点を観察した。

4 非漢字系日本語学習者の読解における辞書使用

4.1 読解環境の整備：学習者が使った辞書ツール

学習者は文章を読み始める前に多くの選択を行うが、一部の選択肢に気づかない学習者もいる。まず、電子媒体の読み物なら、それをコンピュータの画面で読むか、プリントアウトして読むか、両方の媒体を使うかを選ぶ。紙媒体の読み物なら、それをそのまま読む学生がほとんどだが、スキャナーや携帯電話のカメラで読み込み、OCR処理を施しデジタル化して読むという選択肢もある。

次に、辞書として、書籍を使うか、電子辞書専用機（IC電子辞書）、パソコンやタブレットなどでアクセスできるオンライン辞書や一般的な検索エンジン、携帯電話にダウンロードするアプリケーション、電子書籍端末に付属する辞書機能、ゲーム機に搭載できる辞書など、どの媒体を用意するかを選び、辞書を入手する。

今回の調査では、学習者はさまざまな読解環境を整えていた。それぞれの学習者が使ったツールの種類と数は表2に示す。

表2 学習者が使った辞書ツールの数

母語	レベル	オンライン辞書	アプリ	電子辞書	書籍	合計
利用者数	29 (100%)	22 (76%)	10 (35%)	7 (24%)	1 (3%)	
スロベニア語	上	2		1		3
スロベニア語	上	2				2
スロベニア語	上		1	3		4
スロベニア語	上	1		1		2
スロベニア語 +マケドニア語	上	5				5
フランス語	上			1		1
クロアチア語	中					0
スロベニア語	中	3	2			5
スロベニア語	中	2				2
スロベニア語	中	5				5
スロベニア語	中	2				2
スロベニア語	中		2			2
スロベニア語	中	6				6
スロベニア語	中	2				2
スロベニア語	中	1	2			3
スロベニア語	中	1	1			2
ドイツ語	中	3	2			5
スペイン語	中	4				4

スペイン語 +カタルーニャ語	中	1		2		3
フランス語	初	4				4
フランス語	初	1	1			2
フランス語	初			2		2
フランス語	初		1		1	2
フランス語+英語	初	4				4
ドイツ語	初		2			2
スペイン語	初	2				2
スペイン語	初	2				2
スペイン語 +カタルーニャ語	初	3				3
スペイン語 +カタルーニャ語	初	3	1	2		6

表2から分かるように、調査対象29人の内、単独機器の電子辞書を持っている学習者は6人で、書籍の辞書を調査に持参したのは1人しかおらず、この学習者も1時間半の調査のうち1度だけ書籍の辞書を確認し、それ以外の検索はすべて携帯電話のアプリで行った。携帯電話のアプリを使ったのが12人で、オンライン辞書を使ったのが18人で最も多かった。

他の日本国内外の調査でもここ数年の間、書籍と電子辞書の使用が減り、オンライン辞書とアプリの使用が増えたと報告されている。例えば、鈴木他(2019)が2011年と2017年に日本国内の留学生を対象に実施した辞書使用アンケート調査では、単独機の電子辞書を使う留学生が2011年に7割だったのに対し、2017年には25%に減り、アプリ利用者は2011年に2割だったのに対し、2017年には77%に上っていたという。日本国内の調査を実施したToyoda(2016)も、Hmeljak Sangawa(2018)が報告するスロベニアでの調査も同じような傾向が現れており、書籍の利用が極めて稀で、アプリの利用が増加していることが伺える。

辞書及び検索ツールの数に関しては、電子辞書の対訳辞典しか使わなかった1人の上級学習者を除いて、今回の調査の学習者は2つ以上のツールを使い分けていた。多い場合は、携帯電話のアプリもパソコンのオンライン辞書も電子辞書専用機の国語・漢和・和英辞典も同時に使い分ける学習者もあり、使用ツールの平均数は3.1であった。

読解環境として用意する辞書の本数は、日本語レベルによる傾向は見られず、個人的な学習スタイルによる選択のようだ。

例えば、同じ上級レベルの学習者でも、漢字の読みを部首や画数から調べることなく、電子辞書の和仏辞典しか使わなかった学習者もいれば、携帯電話のアプリの和英・漢英辞典の他に電子辞書の和英、国語、漢和辞典を使い分ける上級学習者、そして5つのオンライン辞書を使い、同じ言葉を納得できるまで複数の辞書で確認した学習者もいた。初級レベルの学習者の中にも、オンラインの和英辞典とグーグル検索しか使わなかった学習者もいれば、携帯電話のアプリの和英・漢英辞典、オンラインの辞書ポータルサイトと電子辞書の和英、和英辞典を同時に使う学習者もいた。

今回の調査の学習者が使ったツールの名前は表3に示す。

表3 学習者が使った辞書ツール名

辞書の種類		使用者数	辞書名と使用者数
パソコン	和英オンライン辞書	19名	Jisho 15名, Google Translate 9名, Weblio 5名, Wikipedia 3名, コトバンク 2名, WordReference 1名, goo 辞書 1名
	英語・母語オンライン辞書	10名	PONS (英語・スロベニア語辞書) 7名, Rječnik (英語・クロアチア語辞書) 1名, WordReference (英語・フランス語辞書) 1名
	ウェブ検索	8名	Google 8名, Ask 1名
	日本語・母語オンライン辞書	4名	日仏辞典 2名, 瑠偉オンライン(和西)辞典 1名, 和独辞典 1名
	ブラウザ・アドオン	2名	Firefox Rikaichan 1名, Chrome Rikaikun 1名
携帯アプリ	和英辞書	14名	imiwa? 4名, JED 4名, Takoboto 2名, Jisho 1名, Shirabe 2名, Yomiwa 1名
	英英辞書	1名	Merriam Webster 1名
電子辞書	和英辞書	6名	新和英辞典 4名, ジーニアス和英辞典 1名
	日本語・母語辞書	3名	白水社和西辞典 2名, 旺文社和仏辞典 1名
	国語辞書	3名	大辞泉 2名, 明鏡国語辞典 1名
	漢和辞書	2名	漢語林 1名
ゲーム機	ゲーム機搭載辞書	1名	DS 楽引辞典 1名
書籍	書籍辞書	1名	Dictionnaire Assimil Kernerman 和仏辞典 1名

表3から分かるように、和英のオンライン辞書やアプリを使う学習者が多い。スロベニア語やクロアチア語話者の場合は、読解に利用できる日本語と自分の母語の対訳辞書が存在しないため和英辞典を使わざるを得ないが、フランス語話者やスペイン語話者の場合は、日本語と母語の対訳辞典が書籍としても電子辞書とオンライン辞書としても存在するのに、和英辞典も使う学習者が多い。

辞書を選んだ理由としては、インターフェースが使いやすいこと、入手しやすいこと、値段が安いこと、友達にすすめられたこと、確実な情報を得るために複数の辞書で確認したいことなどの理由があげられた。

このように、多くの学習者は知人や友人の意見や、インターネット上の評価や情報を参考にツールを選び、入手する。そこで優先される要因は、ツールの使いやすさと、入手のしやすさ及び値段のようである。

4.2 検索における工夫と困難点

読解環境の整備ができて、その道具を使いこなせば、必要な情報に早くたどり着くはずである。フメリヤク（1996）の調査で書籍の辞書を使っていた学習者は、読めない漢字を漢和辞典で探し、漢字の読みがわかったら和英辞典で英訳を探し意味を理解しようとしていたが、読解に費やした時間の5割から8割、索引を通して和英の見出しにたどり着く作業に使っていた。今回の調査で電子媒体の読み物も辞書も同じコンピュータの画面に用意した学習者は、1クリックで瞬時に情報にたどり着くことができた。そこで学習者によってはさまざまな工夫が見られた一方、検索が困難だったケースもあった。

【検索対象の特定】分かれ書きされていない日本語の文字列から、特にひらがなだけの長い文字列や漢字熟語を区切り、検索対象を特定することは困難である。例えば、「国旗掲揚式」という文字列を読んだ学習者は、最初の「国」を見落とし、「はた」を入力し「旗」に変換し、「掲」を手書き文字認識画面に指でなぞり入力し、「旗掲」を検索したが、そのような見出し語がなかったため「掲」も指でなぞり入力し「旗掲掲」を検索し、これも何も見つからなかったため「旗」を消し「掲掲」を検索し、やっと見出しを見つけて、そこから5字熟語を理解し、文全体を解釈できた。

【検索すべきかどうかの判断】知らない漢字や単語が多く、調べる時間があまりないときは、どの未知語が重要で調べる価値があるか、どれを調べなくていいかという判断も重要である。しかし、多義語の一つの意味を知っている学習者は、その意味が文脈に合わなくても、自分が知っていると思っている語は辞書で調べない傾向がある。たとえば、「値段をそらで言える」という文を読んだ学習者は、「そら」という単語が「空、天」という意味と「空の空間」という意味で使われると記憶しており、この意味が文脈に合わない判断したが、他にも調べた単語が多く、「そらで」を飛ばしてもいいと判断し、文全体の大体の意味を理解した。

【辞書形（検索する見出し）の決定】「Jisho」や「Weblio」など、多くのオンライン辞書では、複数の単語からなる文字列や否定形などの活用形を入力しても、辞書ポータルが形態素解析を行い、文字列を単語に区切って、その単語の見出しを提示する仕組みになっているので、どの文字を切り取って、どの単位を検索すればいいかが分からないとき、文をまるごとコピーし、オンライン辞書にペーストする学習者もいる。しかし、それでも検索に成功しないことがある。例えば、ある学習者は「切り離せない」という表現を検索するために、可能形だということに気づかず、否定形だけを肯定形に変換し「切り離せる」という単語を辞書で調べたが、見つからなかった。

【辞書、検索方法を選択】さまざまな辞書ツールを巧みに使い分ける学習者もいるが、自分が持っている辞書のさまざまな検索方法を知らないため苦勞する学習者もいる。例えば、電子媒体で手に入った文章を読んでメモするためにプリントアウトし、読めない漢字は、画面上の文章からコピーしてオンライン辞書にペーストするのではなく、画数を数えたり部首を探したりして、必要以上に検索に時間をかけた学習者もいた。

【見出し検索の実行】検索においてさまざまな困難のため検索が成功しないことがある。たとえば、「恋愛対象」という熟語の英訳を調べるために「Jisho」にローマ字で「renaitaisho」と入力した学習者がいる。それで何も見つからなかったため、入力した文字と読み物の言葉を見比べて違いに気づき、日本語入力モードに切り替え、「れないたいしょ」、「れんあいたいしょ」、「れんあいたいしょう」といくつかの入力を試して「恋愛対象」への変換にやっと成功した。

【辞書項目からの情報抽出】見つかった見出し語に対して語義及び訳語が複数ある場合は、その情報の中から文脈に合う情報を見つけ出す必要があるが、語義及び訳語が多い場合は最初の語義・訳語だけを確認し、文脈に合わなくても全ての語義を確認しない学習者もいた。例えば、「という点にその価値は集約されています」という文を読んだ学習者は、「集約」という単語を知らなかったので調べたが、表示された英訳のうち、この文脈に合わない最初の英訳「intensiveness」だけを読んで、二つ目の英訳「summarizing」という単語を読まなかったため、文全体があまり理解できなかった。

【検索結果を文脈に適用】英語などの場合は、綴りが間違っていたら何も見つからないことが多いが、日本語は発音の似ている言葉が多く、誤入力した文字列が見出し語として存在する可能性が高いため、学習者が探している単語とは無関係な見出し語が辞書に表示されることがある。そこで、文字の違いに気づかない学習者は、検索結果を無理に文脈に当てはめようとすることがある。例えば、「取り組まなければならない問題」という表現を読んだ上級学習者は、「取り組む」という単語の意味を文脈から推測できて、その発音も大体覚えていたが、確認するために辞書で調べようとして、「とりくむ」ではなく「とりこむ」と入力した。それで、予想していた「取り組む」という意味の英訳が表示されなかったので戸惑った。

5 この調査の限界と今後の課題

今回行った調査では、さまざまな辞書使用を観察し、検索プロセスの段階とそれぞれの困難点を確認・概観したが、どの段階がもっとも難しく、どの困難点で学習者がもっとも頻繁につまずくかを明らかにするために、読解調査のデータを量的にも分析できるように、表1に示したような項目のアノテーションを施す必要がある。また、桑原 (2019)のような日本語学習者の辞書使用のより詳しいデータ収集と分析が必要であると同時に、そのような調査結果に基づいた、刊行予定の野田他 (2020) のような辞書指導案も望まれる。

謝 辞

本研究は JSPS 科研費 18F18808 の助成を受け、国立国語研究所のプロジェクト「日本語学習者のコミュニケーションの多角的解明」および JSPS 科研費 23320107 と 15H01884 による成果『日本語非母語話者の読解コーパス』を利用した。

文 献

- Hartmann, R.R.K. (1989) *Sociology of the dictionary user: Hypotheses and empirical studies*.
 Hausmann, Franz Josef, Reichmann, Oskar, Wiegand, Herbert Ernst, and Zgusta, Ladislav (ed.):
Dictionaries: An International Encyclopedia of Lexicography, Berlin: De Gruyter. pp.102-111.
 フメリヤク・クリスティーナ (1996) 「日本語学習者の辞書使用における実態調査」『日本語教育方法研究会誌』 3:1, pp.18-19 [<http://ci.nii.ac.jp/naid/110009496640>]
 Hmeljak Sangawa, Kristina (2018) *Japanese language teaching at tertiary level in Slovenia: Past experiences, future perspectives*, *Acta Linguistica Asiatica* 8:1, pp.51-64
 [<https://revije.ff.uni-lj.si/ala/article/view/7702>]
 桑原陽子 (2019) 「非漢字系中級学習者の論文読解における辞書使用のありかたの変化」『国際教育交流研究』 3, pp.1-17.
 [<https://www.u-fukui.ac.jp/wp/wp-content/uploads/Yoko-Kuwabara-1.pdf>]

- 野田尚史 (2019) 「読んで理解する過程の解明—「読解コーパス」の開発—」, 野田尚史・迫田久美子 (編) 『学習者コーパスと日本語教育研究』 東京: くろしお出版. pp.23-42.
- 野田尚史・村田裕美子・中島晶子・白石実 (2020 予定) 「ヨーロッパの日本語学習者の辞書使用の問題点とその指導」『ヨーロッパ日本語教育』 23. ヨーロッパ日本語教師会
- 鈴木智美 (2012) 「留学生の辞書使用についての実態調査—東京外国語大学で学ぶ留学生へのアンケート調査の結果と分析—」『東京外国語大学留学生日本語教育センター論集』 38, pp.1-21 [<http://hdl.handle.net/10108/70119>]
- 鈴木智美・中村彰・清水由貴子・渋谷博子 (2019) 「ICT時代の日本語学習者はどのような学習ツールを使っているか: 日本語教師を対象としたワークショップ実施報告」『東京外国語大学留学生日本語教育センター論集』 45, pp. 239-256 [<http://repository.tufs.ac.jp/handle/10108/92986>]
- Toyoda, Etsuko (2016) Usage and efficacy of electronic dictionaries for a language without word boundaries, *The EUROCALL Review* 24:2, pp. 13-23 [<https://doi.org/10.4995/eurocall.2016.5662>]

関連 URL

『日本語非母語話者の読解コーパス』 <https://www2.ninjal.ac.jp/jsl-rikai/dokkai/>