

世界の日本語学習者の語彙検索行動の概観

—「語彙検索行動情報データセット」の公開—

佐野 彩子・吉 甜・石黒 圭

国立国語研究所 研究系

責任著者 石黒 圭 ishigurokei@ninjal.ac.jp

要旨

近年、日本語学習において、学習者はスマホをはじめとするデジタル端末を使って言葉の意味・用法等を調べるのが一般的である。しかし、学習者が日常的にどのように言葉を調べているか、その実態は明らかになっていない。そこで、国立国語研究所の共同研究プロジェクト「学習者の辞書資源使用の実態調査」チームでは、世界の8つの国・地域で学ぶ日本語学習者110名を調査対象として語彙検索行動の調査を行った。本稿では、その調査から得られたデータのうち、学習者が自身の語彙検索行動を記録した内容を集計した「語彙検索行動情報データセット」の概要、およびその集計結果を報告する。このデータセットには、調査期間である連続する7日間に、学習者がどのような言葉をどのように調べたのかという語彙検索行動にかんする5W1Hのほか、検索目的の達成の有無、その理由にかんする自己申告などが含まれており、このデータセットを用いることで、学習者の日々の辞書検索行動の実態や特徴をつかむことが可能になる*。

キーワード： スマホ、辞書使用、意味・用法、語種、国語辞典

1. はじめに

昨今の情報技術の進化に伴い、日本語学習者の学習環境も大きな変化を遂げている。特に語学学習にとって欠かすことのできない「言葉を調べる」という行為についても、スマホのアプリやインターネットのウェブサイト、また ChatGPT をはじめとする大規模言語モデルをベースとした生成 AI の利用など、その種類も方法も無数に存在し、実態の全容を把握することは容易ではない。学習者向けの辞書については、東京外国語大学の鈴木氏を中心とする研究（鈴木ほか 2018, 中村ほか 2020）が挙げられ、学習者目線に立った辞書ツールの整備が不十分であることが指摘されている（鈴木ほか 2020）。また、実態調査では、調査者から学習者にたいして提示された課題をこなす過程で行われた作文調査（鈴木・高野 2015）、読解調査（桑原 2019, 野田ほか 2020）がある。これらの先行研究の知見をふまえ、石黒ほか（2023）は、調査者による状況設定のない、できる限り自然な状態での語彙検索行動を観察することを試みた。本稿は、この研究において学習者から提供を受けた調査データの1つである「語彙検索行動情報シート」について、その目的

* 本稿は、科研費（JPK21K18375）および国立国語研究所機関拠点型基幹研究プロジェクト「学習者の辞書資源使用の実態調査」の研究成果の一部である。なお「語彙検索行動情報データセット」は、以下に公開している。https://www2.ninjal.ac.jp/jll/word_search/survey-data/information-sheet.html

や内容を説明するとともに、調査対象者 110 名のデータを統合し、複数の情報について追加のタグを付与し作成した「語彙検索行動情報データセット」の集計結果から浮かび上がる学習者の検索行動の実態や特徴を明らかにしたものである。

2. 研究の概要

はじめに、本研究の概要について述べる。本研究は、世界 8 つの国・地域で学ぶ日本語学習者 110 名を調査対象としている。調査時期は 2022 年 1 月から 12 月で、調査期間は学習者の任意の連続する 7 日間（授業のある平日と授業のない週末を含む）であった。

調査対象者には、学習歴などの背景情報、辞書の利用状況、日本語の使用状況にかんする情報確認のためのフェイスシートを記入してもらった後、日本語力を知るためのレベルチェックテストを受験してもらった。当研究室で作成した語彙・文法テスト（市江ほか 2023）と筑波日本語テスト集「TTBJ」に収録されている SPOT の 2 つの日本語テストからレベルを測定している。

続いて、調査期間中に毎日、スマホなどで辞書リソースを使用して言葉を調べたときの様子を画面録画機能を使って検索行動ごとに録画してもらった。同時に、原則として毎日、録画映像に対応する語彙検索行動の 5W1H を「語彙検索情報記録シート」というエクセルのファイルに記入を依頼した。この「語彙検索情報記録シート」110 名分を統合した上で、調査者が複数の項目について追加のタグを付与して作成したのが「語彙検索行動情報データセット」（以下、データセット）である。

3. データセットの目的

データセットの目的は、主に 2 つある。1 つ目の目的は、学習者の使用している辞書リソース¹の正確な把握である。実際に学習者によって使用される辞書リソースは多種多様であるため、録画映像だけを見ても、それが何という辞書アプリであるのか、どのようなウェブサイトであるのか分からない場合も多い。また、学習者の検索行動は、自分で言葉を入力するだけでなく、検索結果の画面に表示された語のハイパーリンクから新たなページに瞬時にとんだり、検索エンジンに表示された結果一覧の 1 つをクリックして別サイトに移動するような検索行動をとるケースも多い。検索行動の 1 つ 1 つをエクセルシート 1 行ごとに自己申告してもらうことで、学習者が実際に「どこにどうやってアクセスしたのか」という、映像からだけでは読み取ることが難しい辞書リソースとその経路をより正確に把握することができる。

2 つ目の目的は、学習者自身の行動や目的の詳細な把握である。学習者はなぜその言葉を検索したのか、調べたい内容は語の読み方なのか、意味なのか、さらには、どうして検索目的が達成されなかったのかという理由を自己申告してもらうことで、個々の検索行動の意図や背景の詳細が分かる。コロナ禍で対面でのフォローアップインタビュー等も実施が難しかったという経緯もあり、このような情報の提供がフォローアップインタビューを補完する内容となった面もある。

¹ 本研究では、「辞書リソース」という語を語彙検索に使用可能なリソースすべて（辞書アプリ、インターネット上の辞書、検索エンジン、機械翻訳、Wikipedia、生成 AI 等）として論じる。

なお、このデータセットは学習者の自己申告ということで、原則として学習者の申告内容をそのまま掲載している。データセット提出にかんして、学生の所属大学の先生方からの指摘に基づいて、データの不備などについて修正し再提出した学生もいたと聞いている。ただし、国立国語研究所に調査データが提出された後、調査者が行ったのは、次の①～③のみである。

- ① 中国語、韓国語、ベトナム語の記述の翻訳
- ② 書体の統一
- ③ タグ付けのための表記の統一（例えば、「大学寮」と「宿舎」、「家」と「自宅」等の表記）とタグの付与

そのため、学生のデータには無回答や申告漏れが存在しているが、原則として学生の申告のままに掲載していることを記しておく。

4. データセットの概要

データセットは、エクセルファイルの 4,713 行、25 列で構成されている。1 行目には、「語彙検索行動情報シート」で申告を依頼した項目、および調査者が付与したタグ（★マークがついている）が入っている。項目の一覧は以下の表 1 である。2 行目以降は、学習者の回答が ID 順に掲載されている。原則として、1 つの言葉を検索するのに 1 つの辞書リソースにアクセスした場合、同じ行に情報を回答し、検索語や辞書リソースを変更した場合には行を変更して回答してもらっている。

表 1 データセット 1 行目の項目一覧

A 列	: ① 大学 ID
B 列	: ② 個人 ID
C 列	: ③ ★追加タグ★ 日本語レベル
D 列	: ④ 調べたい言葉（語・句・文）
E 列	: ⑤ ★追加タグ★ (D 列の) 調べたい言葉の語種
F 列	: ⑥ 実際に調べたかどうか（調べた○／調べていない×）
G 列	: ⑦ ★追加タグ★ 行動種類（単独検索／連続検索）
H 列	: ⑧ 調べた日（調べたいと思った日）
I 列	: ⑨ 調べた場所
J 列	: ⑩ どこで見た／聞いた言葉か
K 列	: ⑪ 知りたい内容（読み、発音、意味、用法など）
L 列	: ⑫ ★追加タグ★ (K 列の) 知りたい内容
M 列	: ⑬ (検索の) 具体的な状況
N 列	: ⑭ 入力内容
O 列	: ⑮ ★追加タグ★ (N 列の) 入力内容の言語
P 列	: ⑯ 使用ツール（スマホ、タブレット、PC など）

Q 列：⑰ 使用したアプリ名，辞書名，サイト名など
 R 列：⑱ ★追加タグ★ (Q 列の) 使用した辞書リソース
 S 列：⑲ Web または App
 T 列：⑳ 入力方法
 U 列：㉑ T 列で「その他」を選んだ場合，記入
 V 列：㉒ ★追加タグ★ (T 列，U 列の) 入力方法
 W 列：㉓ 目的が達成できたか
 X 列：㉔ W 列で「達成できた」以外を選択しなかった場合の理由
 Y 列：㉕ F 列が×の場合，実際に調べなかった理由

また，データセットの項目に含まれる情報は，おおむね，以下の 5 つで構成されている。

- (1) 学習者個人にかんする情報（所属や ID，日本語レベル）
A 列，B 列，C 列
- (2) 検索対象にかんする情報（調べたいと思った言葉，実際に調べたかどうか等）
D 列，E 列，F 列，G 列
- (3) 検索状況にかんする情報（検索を行った日，場所，どこで見た言葉か等）
H 列，I 列，J 列，K 列，L 列，M 列
- (4) 検索内容にかんする情報（入力した内容や，使用したデバイス，辞書リソース等）
N 列，O 列，P 列，Q 列，R 列，S 列，T 列，U 列，V 列
- (5) 検索行動にかんする情報（検索行動によって目的が達成できたか等）
W 列，X 列，Y 列

なお，調査者が付与した追加タグにかんする説明は，別に公開している「追加で付与したタグにかんする補足説明」²に判定方法や基準の詳細が示されているので合わせて参照されたい。

5. データセットの主な集計結果

以下では，上記 A 列から Y 列で見た，データセットの項目の情報の集計結果から，本調査の概要や学習者の辞書検索行動の特徴を明らかにしていきたい。

5.1 所属地域（A 列）と検索数

学習者の所属地域毎の言葉の検索数は表 2 の通りである。

² 「追加で付与したタグにかんする補足資料」は以下に掲載している。
https://www2.ninjal.ac.jp/jll/word_search/survey-data/information-sheet.html

表2 所属地域と検索数

	報告数		調査対象 人数	1日の検索数 (1名あたり)
		うち検索数		
中国語簡体字圏A大学	1,300	1,248	28	6.4
中国語繁体字圏B大学	362	359	20	2.6
中国語繁体字圏C大学	473	470	11	6.1
韓国語圏D大学	428	427	11	5.5
韓国語圏E大学	409	402	11	5.2
ベトナム語圏F大学	1,053	1,036	10	14.8
英語圏G大学	100	96	3	4.6
ドイツ語圏H大学	101	99	3	4.7
日本国内I大学	486	486	13	5.3
総計	4,712	4,623	110	6.0

学習者の1日あたりの平均検索数は6.0語となった。ベトナム語圏F大学の検索数は14.8語と群を抜いて高く、中国語繁体字圏B大学は2.6語と少なかった。日本語学習を主専攻としているかどうか、調査時期の授業課題、辞書リソースの使い方の習慣など様々な要因が検索語数の開きの要因として考えられる。ただし、上記2大学を除く大学では、1日あたりおよそ4語から6語の域に入っている。なお、実際に調べようと思ったが、調べなかった理由として、「後から意味や読み方を（調べなくても）思い出したり、推測したりできた」「先生が授業で教えてくれた」等の理由も挙げられている。

5.2 所属地域（A列）と日本語レベル（C列）

調査参加大学ごとの調査対象人数と日本語レベル、またフェイスシートで確認した学習者の母語の内訳を表3に示す。

表3 調査参加者数と日本語レベル

	日本語レベル					総計	母語内訳
	初級	中級	中上級	上級			
中国語簡体字圏A大学	2	9	12	5	28	中国語27, モンゴル語1	
中国語繁体字圏B大学	14	3	3		20	中国語9, 広東語10, 英語1	
中国語繁体字圏C大学			7	4	11	中国語11,	
韓国語圏D大学	2	5	4		11	韓国語11	
韓国語圏E大学			5	6	11	韓国語11	
ベトナム語圏F大学	5	5			10	ベトナム語10	
英語圏G大学	1	2			3	英語3	
ドイツ語圏H大学	2	1			3	ドイツ語3	
日本国内I大学			11	2	13	中国語11, ベトナム語1, ドイツ語1	
総計	26	25	42	17	110		

海外の大学については、各大学の教員にご自身の担当するクラスを中心に調査対象学生を募っていただいた経緯もあり、おおよそ大学ごとに日本語レベルがまとめられる。中国語圏では、中上級者を中心とする A 大学、C 大学、初級者を中心とする B 大学、韓国語圏では、中級者を中心とする D 大学、上級者を中心とする E 大学、ベトナム語圏、英語圏、ドイツ語圏では、いずれも初級者から中級者を中心としている。日本国内の大学では、中上級者以上が最も多い。調査対象全体では、中上級者が 42 名と最も多く、次いで初級者 26 名、中級者が 25 名と続き、最後に上級者が 17 名となり、中上級を中心にしてすべての日本語レベルをカバーしていることが分かる。また、学習者の母語と照らし合わせてみると、中国語が 47 名（4 地域にわたる）と最も多いものの、所属は中国語圏の大学であるが、母語は広東語と申告した学生が 10 名あった他、英語 1 名、モンゴル語 1 名が含まれる。また、日本国内の大学に所属している学生にも中国語が母語の学生以外にもベトナム語 1 名、ドイツ語 1 名が含まれ、調査地域、学習者母語ともバラエティに富んだ構成になっていることが分かる。

5.3 「調べたい言葉」（D 列）と「調べたい言葉の語種」（E 列）

D 列の「調べたい言葉」は、学習者が検索を行う際、調べようと思った言葉で、実際に辞書リソースに「入力した言葉」とは必ずしも一致しない。例えば、学習者の母語で「**」は日本語で何というのかを調べたいときには、「調べたい言葉」に学習者の母語の「**」が記入される。E 列では、D 列に記入された「調べたい言葉」の語種を判定し、語種のタグを付与した。語種の判定基準については、「追加で付与したタグにかんする補足資料」を参照されたい。E 列「調べたい言葉の語種」の集計結果を C 列の学習者の日本語レベルごとに集計した結果を、以下の表 4 に示す。各語種の合計数を、各日本語レベルの合計人数で割った 1 名あたりの「調べたい語」の語数を語種別に表示している。

表 4 日本語学習者レベル別の調べたい語の語種（1 名あたり）

	和語	漢語	外来語	混種語	固有名詞	外国語	対象外等
初級	10.4	13.7	1.1	1.4	0.3	0.6	1.5
中級	14.5	20.7	2.0	1.7	0.5	0.7	1.8
中上級	11.9	11.6	3.0	1.5	0.7	0.3	2.1
上級	12.1	11.0	4.8	3.5	0.7	0.1	2.5
総計	48.9	57.0	11.0	8.1	2.2	1.7	7.9

表 4 から、和語と漢語は中級を除くすべてのレベルで 1 週間におよそ 1 人あたり 10 語程度検索されていることがわかる。中級の学習者では、和語が 14.5 語、漢語が 20.7 語と語数が多かった。また、日本語レベルが上がるにつれて、外来語、混種語の言葉が増加していることもわかる。上級レベルになるにつれて、より抽象性の高い語や、語構成が複雑な語が検索されていることが挙げられるだろう。さらに「対象外等」には、「百聞は一見に如かず」「情けは人のためになら

ず」のような語レベルでの判定がつかない慣用句等の表現が含まれていた。上級レベルでは、このように語種判定ができない表現レベルの検索も多いという状況もうかがえる。

5.4 行動種類 (G 列)

学習者の語彙検索行動は、1回の入力で1つのリソースにアクセスして検索を行う単独検索と、入力内容やリソースを変更して続けて検索を行う連続検索に分類される。日本語レベル別に行動の種類を分類した集計は表 5、また 1 名あたりの検索数を示したのが表 6 である。

表 5 日本語レベル別の単独検索，連続検索数

	単独	連続	2回	3回	4回	5回	6回	7回	8回	9回	14回
初級	642	103	63	17	13	5	3	0	2	0	0
中級	867	109	72	21	8	5	0	0	1	1	1
中上級	1,059	235	163	46	11	6	5	2	1	1	0
上級	396	183	107	51	16	8	1	0	0	0	0
総計	2,964	630	405	135	48	24	9	2	4	2	1

表 6 日本語レベル別の単独検索，連続検索数 (1 名あたり)

	単独	連続
初級	24.7	4.0
中級	34.7	4.4
中上級	25.2	5.6
上級	23.3	10.8
総計	108	25

調査全体数のうち、単独検索は 2,964 件 (82.5%)、連続検索は 630 件 (17.5%) である。なお連続検索の場合は、2 回 (連続して 2 つの辞書リソースを利用して検索を行う) が最頻値となり 405 回、連続検索の最長は 14 回の検索であった。また表 6 からは、日本語レベルが上がるにつれて、連続検索が増え、とくに上級では連続検索が一気に多くなることが分かる。学習者が連続検索を行う理由 (喬 2024)、日本語力の伸びの差と検索行動の違い (袁 2024) において詳細に分析されているが、日本語力が上がるにつれて、類似する語や表現の意味、用法の微妙な違いについて調べようとしたり、複合語、新語、専門用語、コロケーション等まで調べ、自分の理解を確認したり、より具体的なイメージをつかんで知識を広げ、理解を深めようとする姿勢が見られることが背景にある。

5.5 検索場所 (I 列)

学習者が語彙検索行動を行った検索場所を集計した結果は、以下の表 7 である。学習者の日本語のレベルごとに集計すると、どのレベルの学生においても自宅での検索数が学校や外（学校以外）よりも多い。全体のおよそ 76.7%が自宅で検索し、学校 16.0%、外（学校外） 6.4%となった。調査自体がコロナ禍の自宅待機中に行われたことも影響していると推察される。なお、調査対象者の中には寮生活をしている学生もいるが、寮は「学校」に含まれている。また、「その他」には、オンラインでの日本語練習等も含まれている。

表 7 学習者のレベル別の検索場所と検索件数

	自宅	学校	外（学校以外）	その他
初級	521	187	36	6
中級	839	182	31	0
中上級	988	166	140	28
上級	496	59	32	0
総計	2,844	594	239	34
割合	76.6%	16.0%	6.4%	0.9%

5.6 知りたい内容 (L 列)

学習者の知りたい内容を、レベルごとに集計し、総計が 20 以上となった項目を表示した結果が以下の表 8 である。

表 8 知りたい内容と日本語レベル

	初級	中級	中上級	上級	総計	総計割合
意味	383	502	611	338	1,834	38.9%
意味、読み方	206	315	311	47	879	18.7%
読み方	32	127	184	58	401	8.5%
意味、用法	39	99	59	73	270	5.7%
用法	15	22	50	36	123	2.6%
日本語の言い方	25	10	37	17	89	1.9%
表記（漢字_書き方）	12	2	22	1	37	0.8%
読み方、用法	17	3	7	3	30	0.6%
表記（漢字）	4	8	3	12	27	0.6%
意味、読み方、用法	4	7	5	10	26	0.6%
母語の言い方	22	0	2	1	25	0.5%
意味、語源	0	0	2	20	22	0.5%
該当なし	154	111	299	214	778	16.5%

知りたい内容では、すべてのレベルの学習者において、日本語の「意味」を調べる検索が最も多く全体の38.9%を占めた。また、上級を除く初級、中級、中上級では、「意味、読み方」が次に多かった。一方、上級では「意味、用法」を検索が「意味」に次いで多かった。「意味」「意味、読み方」「意味、用法」と意味を含む検索内容の上位3項目の割合を合計すると63.3%を占め、検索内容のおよそ6割が語の意味を調べることを目的とした検索であったことがわかる。

なお、L列を集計すると「該当なし」が総計778(16.5%)あった。これは、K列が無回答だったもので、うち766が連続検索の過程での空欄であった。

5.7 入力語の言語 (O列)

学習者が検索をする際、入力した言語について集計した結果は、表9である。全体のおよそ89%は、日本語を入力しており、日本語と外国語入力を組み合わせた形(例えば、「水商売 meaning」(日本語+英語)「下班 日本語」(中国語+日本語))を合わせるとおよそ9割が、入力に日本語が含まれている。これは、前節の「検索内容」で「意味」を調べることを目的とした検索が多かったこととも対応している。また、後節に述べる「辞書リソース」には、中国語圏や韓国語圏の中級上級、上級を中心とする学生の中に、国語辞典等の日日辞典の利用が見られており、母語を介さずに日本語の意味や用法を直接日本語のまま理解しようとしている状況とも合致する結果である。

なお、O列を集計すると「該当無し」が総計186(3.9%)ある。これもO列が空欄だったもので、連続検索の一部等で入力行為そのものが不要だったもの(ハイパーリンクへのジャンプ)97、実際に調べなかったもの89が含まれる。

表9 所属大学別の入力語の件数

	合計	割合	中国語	中国語	中国語	韓国語圏	韓国語圏	ベトナム語圏	英語圏	ドイツ語圏	日本国内
			簡体字圏 A大学	繁体字圏 B大学	繁体字圏 C大学	D大学	E大学	F大学	G大学	H大学	I大学
01_日本語	4,189	88.9%	1,115	325	435	353	349	1,022	75	79	436
02_中国語	122	2.6%	66	14	12						30
03_英語	37	0.8%		2	3	1			21	7	3
04_韓国語	99	2.1%				52	47				
05_ベトナム語	18	0.4%						12			6
06_ドイツ語	12	0.3%								11	1
07_日本語+ベトナム語	1	0.0%						1			
07_日本語+英語	8	0.2%	1	6							1
07_日本語+韓国語	3	0.1%				3					
07_日本語+中国語	7	0.1%		6	1						
08_中国語+日本語	12	0.3%	2	3	7						
09_韓国語+日本語	1	0.0%				1					
10_日本語読み・韓国語表記	17	0.4%				17					
11_該当なし	186	3.9%	116	6	15	1	13	18	4	4	9
総計	4,712	100.0%	1,300	362	473	428	409	1,053	100	101	486

5.8 辞書リソース（R列）

使用された辞書リソースについて、所属大学ごとに使用数を集計したのが、表10以降である。なお、最右列の割合は、各地域の辞書リソース使用総数に対する該当リソースの割合を示す。

表10 中国語圏大学の学習者が使用した辞書リソース（Top5）

	中国語簡体字圏 A大学	中国語繁体字圏 B大学	中国語繁体字圏 C大学	総計	割合
01_MOJi辞書	662	62	115	839	39.3%
04_Google	24	19	127	170	8.0%
06_Weblio辞書	43	15	59	117	5.5%
10_スーパー大辞林	99	11		110	5.2%
09_网易（有道词典）	73	36		109	5.1%

表11 韓国語圏の学習者が使用した辞書リソース（Top5）

	韓国語圏 D大学	韓国語圏 E大学	総計	割合
03_NAVER	384	274	658	78.6%
03_NAVER papago	24	23	47	5.6%
04_Google		36	36	4.3%
05_goo辞書		19	19	2.3%
06_Weblio		19	19	2.3%

表12 ベトナム語圏の学習者が使用した辞書リソース（Top5）

	ベトナム語圏 F大学	割合
02_Mazii	970	92.1%
04_Google	21	2.0%
22_Easy Japanese	9	0.9%
06_Weblio	5	0.5%
53_Jaccent	4	0.4%

表 13 英語圏・ドイツ語圏の学習者が使用した辞書リソース (Top5)

	英語圏 G大学	ドイツ語圏 H大学	総計	割合
27_takoboto		75	75	37.3%
17_Jisho.org	22	15	37	18.4%
23_Nihongo (Serpenti Sei LLC)	33		33	16.4%
13_Japanese (renzo inc)	22		22	10.9%
20_DeepL	10		10	5.0%

表 14 日本国内の学習者が使用した辞書リソース (Top5)

	国内I大学	割合
01_MOJi辞書	194	48.6%
04_Google	83	20.8%
08_沪江 (小D词典)	35	8.8%
09_网易 (有道翻译)	24	6.0%
05_goo辞書	23	5.8%

表 10 から、中国語圏の上位 3 つの辞書リソース、MOJi 辞書、Google (検索エンジン)、Weblio 辞書については 3 つの大学で共通して使用されていることがわかる。また、中国語母語学習者 11 名を含む国内大学の表 14 と比較しても、所属大学の地域を問わず、中国語圏の学習者では、共通した辞書リソースが頻繁に使用されていることもうかがえる。

表 11 では、韓国の学習者の多くが Naver 辞書、及び Naver 社が提供する機械翻訳 papago の使用し、その割合は 8 割を越えることが分かる。中上級者のみを分析した吉ほか (2024) でも、韓国の学習者は、ほぼ Naver 辞書一択であったことを指摘している。さらに、表 12 からは、ベトナム語圏の学習者の多くが Mazii を使用しており、その割合はさらに高い 9 割以上にのぼる。最後に表 13 からは、英語圏、ドイツ語圏の学習者が Jisho.org を共通して使用していることがわかる。調査対象が少なかった英語圏、ドイツ語圏については分析の精度を欠くことは否定できないが、およそ地域ごとに主要な辞書リソースが存在し、幅広いレベルの学習者に共通して使用されていることがわかる。

5.9 入力方法 (V 列)

入力方法を所属大学ごとに集計し、総計が 10 以上のものを表示したのが表 15 である。

表 15 所属大学ごとの入力方法

	中国語 簡体字圏 A大学	中国語 繁体字圏 B大学	中国語 繁体字圏 C大学	韓国語圏 D大学	韓国語圏 E大学	ベトナム語圏 F大学	英語圏 G大学	ドイツ語圏 H大学	日本国内 I大学	総計
01_キーボード入力	895	281	348	223	263	623	62	84	409	3,188
02_手書き入力	219	14	2	52	50	126	30	9	15	517
03_コピペ入力	6	19	24	104	14	200		3	19	389
05_クリック	97	27	61	31	40	69	1		30	356
04_複数の入力方法 (コピペ+キーボード)	12	1	7	1	5				2	28
11_文字列ドラッグ	3	4	14	3	2					26
04_複数の入力方法 (手書き+クリック)				2	20					22
04_複数の入力方法 (手書き+キーボード)	2			3	2	6	3			16
04_複数の入力方法 (キーボード+クリック)			4	1	3	2			3	13
12_一度検索した語の部分削除や部分追加	1		2	1	1	2			6	13

どの地域においても、キーボード入力をもっとも多い点は共通するが、手書き入力、コピペ入力の割合には違いがみられる。中国語圏の中では、簡体字圏 A 大学では手書き入力が圧倒的に多く、コピペ入力が極めて少ない。一方で、特にコピペの割合が高い、韓国語圏 D 大学、ベトナム語圏 F 大学はいずれも学習者のレベルは、初級～中級を中心としている点が共通している。同レベルの中国語圏 B 大学でもコピペは見られるが、他の入力方法と比較すると限定的と言えるだろう。韓国やベトナムのこのレベルの学習者にとっては、書く負担が大きい「手書き入力」よりも、できる限り「コピペ」を利用し手軽に検索するという、入力のスピード重視の省エネ志向もうかがえる。

5.10 達成度 (W 列)

最後に、検索行動によって、検索目的が達成されたかどうかを集計した結果を表 16 に示す。回答は 4 つ（「達成された」「達成された部分もあり、達成できなかった部分もあった」「達成されなかった」「わからない」）から 1 つを選択し回答してもらった。

表 16 国・地域別の達成度

	達成	部分達成	未達成	わからない	総計	達成率	未達成率
中国語簡体字圏A大学	1,123	52	60	40	1,275	88.1%	11.9%
中国語繁体字圏B大学	264	23	40	11	338	78.1%	21.9%
中国語繁体字圏C大学	323	52	76	16	467	69.2%	30.8%
韓国語圏D大学	316	46	47	17	426	74.2%	25.8%
韓国語圏E大学	354	23	13	4	394	89.8%	10.2%
ベトナム語圏F大学	862	17	128	29	1,036	83.2%	16.8%
英語圏G大学	93	3	1		97	95.9%	4.1%
ドイツ語圏H大学	97	2			99	98.0%	2.0%
日本国内I大学	387	11	41	5	444	87.2%	12.8%
総計	3,819	229	406	122	4,576	83.5%	16.5%

表 16 の未達成率（最右列）は、「部分達成」「未達成」「わからない」の回答の合計の割合である。自己申告をもとにした達成率では全体の平均は、達成率 83.5%、未達成率 16.5%であった。達成率が最も高かったのはドイツ語圏 H 大学の 98%、最も低かったのは中国語繁体字圏 C 大学の 69.2%であった。学習者の辞書リソースの使用は、自己申告よれば、おおむね達成されており、およそ 5 件に 1 件程度が、検索目的が達成されないまま終わってしまっていることがわかる。

なお、学習者による自己申告では、達成と申告されているものの、同時に提出された録画映像から、辞書リソースに表示された内容と検索内容が合致していないケースや、学習者が途中で検索を断念してしまったり不自然と思われる行動をとっているケースなど、検索目的が達成されていないと推察される検索行動が含まれていた。別途、国立国語研究所に所属する母語話者が映像を検証する予定である。

6. まとめと課題

データセットの概要および前節の集計結果をまとめ、今後の課題を述べたい。データセットは、本研究の調査データである学習者から提供を受けた「語彙検索行動情報シート」のデータ 110 名分を統合し、調査者によって必要な追加タグを付与した 4,713 行、25 列で構成されたデータセットである。このデータセットでは、同時に提供を受けた録画映像の情報の補足として学習者がどのような辞書リソースを使用して言葉を調べたのかという情報、また学習者がどのような目的を持って検索行動を行い、その結果、目的を達成できたのかどうかという情報が含まれている。

データセットの主な集計結果からは以下の内容が明らかになった。

1) 学習者の所属地域と検索数

学習者の 1 日あたりの平均検索数はおおよそ 6.0 語となった。ベトナム語圏 F 大学の検索数は 14.8 語と群を抜いて高く、中国語繁体字圏 B 大学は 2.6 語と少なかったものの、この 2 大学を除き、平均検索数は 4 語から 6 語の範囲に収まっていた。

2) 日本語レベルと学習者の属性

調査対象者は、中国語圏、韓国語圏、ベトナム語圏、英語圏、ドイツ語圏、および日本国内の大学に所属しており、日本語レベルは初級から上級まで幅広く分布していた。特に中上級者が最も多く、学習者の母語も多様であった。

3) 「調べたい語」の語種の特徴

学習者が検索する語の種類を分析すると、和語・漢語の検索数はほぼ同程度であり、中級者のみやや高い傾向が見られた。また、上級者ほど外来語や混種語を多く検索する傾向があり、より抽象的・複雑な語彙を調べることが多いことが推察される。

4) 検索行動の種類

検索行動は、単独検索と連続検索に分類された。全体の 82.5%が単独検索であり、連続検索は 17.5%にとどまった。ただし、上級者では連続検索の割合が高く、上級者ほどより詳細な情報を求め、語彙力の向上やより正確な運用を目指した語彙検索を行う様子が見られた。

5) 検索場所

検索が行われた場所としては、自宅が全体の76.7%を占め、学校(16.0%)、外(6.4%)と続いた。自宅の割合が比較的高かったのは、調査がコロナ禍に実施されたことも影響していると考えられる。

6) 検索内容

検索の目的は「意味」の確認が最も多く、全体の38.9%を占めた。また、初級・中級・中上級では「意味、読み方」の検索が多く、上級では「意味、用法」の検索が増加する傾向が見られた。「意味」を含む検索内容の上位3項目を合計すると63.3%を占めており、日本語学習において語義の理解が重要であることがわかる。

7) 入力言語

検索時の入力言語は、約89%が日本語を含んでいた。また、日本語と外国語の組み合わせた形も合わせると、およそ9割に日本語が含まれていた。連続検索の一部では、ハイパーリンクを利用したため入力が不要な場合もあった。

8) 使用辞書リソース

学習者が利用した辞書リソースには、地域ごとの傾向が見られた。例えば、中国語圏ではMOJi辞書、Google、Weblioが共通して使用され、韓国語圏ではNaver辞書・Papago、ベトナム語圏ではMazii、英語圏とドイツ語圏ではJisho.orgが主要な辞書として利用されていた。これらの結果から、各地域には主要な辞書リソースが存在し、多くの学習者が共通して使用していることが確認された。

9) 入力方法

キーボード入力が最も多く使用されていたが、手書き入力やコピー入力の使用割合には地域差が見られた。特に、非漢字圏の初級から中級レベルの学習者が多い大学では、書く負担を軽減するためにコピー入力が多く利用されていた。

10) 達成度

自己申告による検索の達成率は83.5%であり、最も高かったのはドイツ語圏H大学(98.0%)、最も低かったのは中国語繁体字圏C大学(69.2%)であった。検索は、おおむね達成されており、およそ5件に1件程度が達成されないまま終わってしまっていることが分かった。

データセットからは、日本語学習者の語彙検索行動には、日本語レベル、母語や所属大学の地域による影響があることがわかる。また、使用する辞書リソースや入力方法にも地域ごとの特徴があった。データセットは、学習者による語彙検索行動の実態を踏まえた辞書リソースの効果的な活用や指導方法を考える上でも、重要なデータを提供するものであると考えられる。

ただし、語彙検索行動では、目的が達成されたと自己申告されているものの、同時に提出された録画映像と突き合わせてみると、映像に映る辞書リソースに表示されている情報と検索内容が

合致していないケースをはじめ、検索目的が達成されていないのではないかと推察される行動も含まれていた。今後、国立国語研究所に所属する母語話者が映像を再検証する予定である。

参照文献

- 石黒圭・佐野彩子・吉甜 (2023) 「世界の日本語学習者の辞書ツール使用事情—スマホによる語彙検索行動の適切な支援のために—」『社会言語科学』26 (1) 5-20.
- 市江愛・吉田暁・石黒圭 (2023) 「日本語教育研究のための「かんたん日本語テスト」の開発—テスト開発経緯と項目分析結果を中心に—」『国際学報』創刊号 19-27. 東京都立大学国際センター
- 袁建華 (2024) 「同じ学習歴の学習者になぜ日本語レベルの差が生じるのか—中国語話者の語彙検索行動を手掛かりに—」『スマホは辞書になりうるか』81-96. 明治書院
- 吉甜・佐野彩子・石黒圭 (2024) 「日本語学習者の辞書リソース利用の拡張—韓国・中国語話者によるスマホを用いた語彙検索行動の実態—」『専門日本語教育研究』26 51-58.
- 喬曉筠 (2024) 「学習者はなぜ連続検索を行うのか—台湾の中級・上級学習者の事例を通して—」『スマホは辞書になりうるか』97-116. 明治書院
- 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—教育実践論文の読解プロセスの事例報告—」『国際教育交流研究』3 1-13. 福井大学国際センター・語学センター
- 鈴木智美・清水由貴子・渋谷博子・中村彰・藤村知子 (2018) 「予備教育課程の国費学部留学生の学習ツール使用状況—2016～2017 年度実施のアンケート調査の結果から見えるスマートフォンアプリの使用目的の多様化と学習スタイルの変化—」『東京外国語大学留学生日本語教育センター論集』44 196-217.
- 鈴木智美・清水由貴子・中村彰・渋谷博子 (2020) 「海外の大学における日本語学習者のツール使用状況の解明—ICT 時代における教師の教育設計リテラシーの向上を目指して—」『日本語・日本学研究』10 23-48. 東京外国語大学国際日本研究センター.
- 鈴木智美・高野愛子 (2015) 「中上級日本語学習者の辞書使用—作文時の辞書使用の詳細調査と文章表現のための辞書使用スキルアップを目指すワークショップ実践報告—」『東京外国語大学留学生日本語教育センター論集』41 137-156.
- 中村彰・鈴木智美・渋谷博子 (2020) 「日本語学習者の学習ツール使用の変遷をその学習歴から探る—初級から超級までの 6 名の留学生へのインタビュー調査に基づいて—」『東京外国語大学国際日本学研究』0 186-196.
- 野田尚史・村田祐美子・中島晶子・白石実 (2020) 「ヨーロッパの日本語学習者の読解における辞書使用の問題点とその指導」『ヨーロッパ日本語教育』24 185-202.

関連 Web サイト

「語彙検索行動情報データセット」 https://www2.ninjal.ac.jp/jll/word_search/survey-data/information-sheet.html (2025年3月27日確認)

「追加で付与したタグにかんする補足資料」 https://www2.ninjal.ac.jp/jll/word_search/survey-data/information-sheet.html (2025年3月27日確認)

An Overview of Vocabulary Search Behavior Among Japanese Language Learners Worldwide – Publication of the "Vocabulary Search Behavior Information Dataset"–

Ayako Sano, Tian Ji , Kei Ishiguro,
National Institute for Japanese Language and Linguistics

Abstract

In recent years, the use of digital devices, including smartphones, has become commonplace among Japanese language learners for looking up word meanings and usage. However, the actual practices of learners in their daily vocabulary searches remain largely unclear. To address this gap, the research team of the collaborative project "Investigation of Learners' Use of Dictionary Resources" at the National Institute for Japanese Language and Linguistics conducted a study on the vocabulary search behavior of 110 Japanese language learners across eight countries and regions worldwide. This paper provides an overview of the "Vocabulary Search Behavior Information Dataset," which compiles self-recorded vocabulary search activities by the learners, as well as a summary of the dataset's findings. The dataset includes details on the vocabulary search behavior over a continuous seven-day period, capturing the "5W1H" (who, what, when, where, why, and how) of each search. Additionally, it records whether the learners successfully obtained their desired information, along with their self-reported reasons for success or failure. By utilizing this dataset, it becomes possible to gain insights into the daily dictionary search behaviors and characteristics of Japanese language learners.

Keywords: smartphone, dictionary use, meaning and usage, word type, Japanese language dictionary