

漢字学習と語彙学習はどうあるべきか ーテスト, コーパス, データベースの分析から考えるー

松下 達彦 (東京大学)

1. はじめに

本稿では、漢字学習と語彙学習のあり方について、主にテストデータ、コーパス、データベースの分析などに基づいて考察する。前半では、漢字知識と語彙知識の関係について考察し、漢字と語彙の頻度レベルの対応なども踏まえて、中級レベルでの語彙学習について提案する。後半では学習者の第1言語知識の日本語の語彙知識への影響について重要な点について取り上げ、中国語 L1, 英語 L1 学習者の日本語語彙学習について提案する。最後にまとめと今後の課題を述べる。

2. 漢字力と語彙力の関係

2.1 漢字知識, 語彙知識の諸側面

一口に漢字力/語彙力と言っても、処理過程での入力と出力のそれぞれが音声、文字、意味のどれか、文脈がどう与えられるかにより、測定対象の知識は異なる。4技能は、音声→意味(聞く)、文字→意味(読む)、意味→音声(話す)、意味→文字(書く)という処理だと整理できるが、実際には語彙や漢字を「話す」テストは少なく、代わりに疑似的な音韻情報として「かな表記」が多用される。「聞く」も同様で、かな表記を漢字で書かせるテストが多い。さらに、漢字は文字が音を直接に表さないため、(疑似)音韻→文字、文字→(疑似)音韻といった形式の測定が行なわれることも多い。例えば「花がさいている。」という文があり、下線部がかなや音声で与えられるといった形式である。第二言語教育では、音声や文字に加えて意味を入力に加える形式も多い。意味の与え方には、a) 定義の辞書的記述(日本語)、b) 定義の辞書的記述(学習者の言語)、c) 対訳(学習者の言語)、d) 類義語、e) 絵、記号、動画、f) 文脈付与、などがある。f) は「水をやり忘れて、花が_____しまった」のように、連語知識や談話構成知識の利用で正答が決まるようにする方法である。上例では「枯れて」「しぼんで」など複数の正答が出るので、選択肢を設けたり、語形の一部を与えたりすることもある。テスト形式により漢字や語の多様な側面が測定されている。

2.2 漢字知識は語彙知識の一部

単漢字の知識と語の知識は「愛」「味」などの1字漢字語では一致するが、「気」「皿」のように、多くの漢字は語構成要素として働き、実際の意味は語にならないと決まらない字も多い。また、「塩味」「甘味」のように異なる読みで語になることも

あり、訓読みだと語だが音読みだと語構成要素にしかない漢字もある。「蒸気」「気分」「亜流」「東亜」のように語の中で実現する単漢字の意味は微妙に異なり、比喩的な意味拡張の程度も異なる。漢字と語という異なる“層”のそれぞれに音声形態、文字表記、意味があり、その三者間でかなり複雑な対応関係があると言える。

いずれにしても漢字知識は語彙知識の一部であり、意味の単位（ひいては学習の単位）としては単漢字より語のほうが重要である。日本の学校教育では伝統的に漢字学習への意識が強いが、意味・用法のより基本的な単位である語を上位概念としてカリキュラムを考え、その中で必要に応じて漢字を学ぶべきである（松下, 2017）。

2.3 漢字テストと語彙テストの得点の一般的な相関

興味深いことに、漢字と語彙の異なる側面を測定するテスト間でも高い相関が出る。酒井（2005）は漢字の読み、書きなどと語彙知識のいくつかの異なる側面のテストを行なったが、概ね 0.7 を超える相関が出ている。松下ほか（2021）は「かな・文脈→漢字」の漢字変換テスト（KCT）と、「語（文字）→意味記述」の語義理解テスト（VSTRJ-50K）という、異なる処理過程の知識を測定する両テストの総得点の相関係数を $r = .82$ ($n=446, p < .01$) と報告している。これらは、形だけを覚えて意味がわからない、あるいはその逆といったことが少なく、両方を覚えることで知識として安定する傾向があることを示すと思われる。ただし「読む」語彙知識と「聞く」語彙知識の発達過程が学習者の第一言語によって異なる傾向のあることに留意する必要がある。

2.4 漢字レベルと語彙レベルの対応

旧日本語能力試験では、3級（現行試験 N4 に対応）を「漢字（300字程度）・語彙（1,500語程度）を習得」したレベルとし、2級（同 N2 に対応）を「漢字（1,000字程度）・語彙（6,000語程度）を習得」したレベルとしていたが、現実には、漢字と語彙の学習レベルはどう対応するであろうか。Matsushita（2012: 240）によれば、書籍約 2,800 万語と「Yahoo 知恵袋」約 500 万語のテキストでは、ひらがな・カタカナ＋漢字頻度上位 300 字の組み合わせで「語のテキストカバー率」は 82.6% になり、ひらがな・カタカナ＋漢字頻度上位 1000 字では 95.1% に達する。既知語数も 6,000 は越え得るので、漢字 1,000 字で中級終了というのはおおむね妥当だと言えよう。

ただ、気をつけなければならないことがいくつかある。

第一に、学習した漢字の組み合わせでできる語をすべて理解できるとは限らない

表1 頻度レベル別および旧日本語能力試験級別の漢字数の分布 (注1)

漢字頻度 順位	旧日本語能力試験級				小計	級外 (注2)	合計
	4級	3級	2級	1級			
1-100	46	36	18	0	100	0	100
101-300	34	58	104	3	199	1	200
301-1,000	23	76	433	166	698	2	700
1,001-2,000	0	12	183	658	853	147	1,000
2,001-6,330	0	0	1	190	191	4,140	4,331
合計	103	181	739	1,017	2,041	4,290	6,331

注1: 松下 (2014) をもとに Matsushita (2012) Table 6-2 と同様の表を再作表した。

注2: 記号「々」も含む。

ということである。高頻度の基礎語「秋」「季節」を知っていれば、低頻度語の「秋季」の意味もほぼ正しく推測できるが、「名」「人」の意味を知ってい

ても「名人」の意味の推測は難しい。漢字語の意味的透明度(単漢字の意味による語意推測レベル)の研究が重要である。

第二に、高頻度の漢字や語彙が常に先に学ばれるべきだとは単純には言い切れない点である。例えば、旧日本語能力試験で1級もしくは級外(すなわち上級レベル)とされる漢字のうち「義」「態」「郎」の3字は頻度上位300に含まれ、172字が上位1,000字に含まれる(表1, 太字部分)。この172字の中には「離」「瞬」「魔」といった筆画の多くて複雑な字も少なくない。頻度は社会的ニーズを反映しており、高頻度の字や語が重要であることは疑いない。したがって、理解と産出を分け、ある段階では理解できれば書けなくてもよい、といった指導も必要であろう。

第三に、高頻度語にも頻度順位の低い漢字が使われるということである。これには二つのタイプがある。一つはある漢字が一つの語に集中して使われる場合で、「距離」の「距」, 「雰囲気」の「雰」などである。このような漢字は語として学習すればよい。もう一つは、漢字で書くこともあるが仮名で書くことも多いという場合で、「駄目」「奴」「嘘」といった日常語のほか、「即ち」「是非」などの接続詞、副詞も多い。これらは理解さえできれば書く際にはひらがなやカタカナで済むであろう。

このように注意すべき点がいくつかあるが、総じて言えば、高頻度漢字を学習すれば高頻度語をカバーできるようになっていくと言える。

2.5 中級の語彙学習をどう構想するか

第二言語教師は語彙学習がいかに大きな負担であるかを真剣に考える必要がある。上節で「ひらがな, カタカナ+漢字頻度上位1,000字」だけで語のカバー率95%に達すると述べたが、95%のカバー率には上位9,446語(語彙素)を要する(Matsushita, 2012: 155)。逆に言えば、ひらがな, カタカナ+漢字頻度上位1,000字の組み合わせからなる語をすべて習得すれば、9,000語以上が習得できるということである。高レベルの読解に必要な語数はかなり多いが、漢字数はそれほどでもないとも言える。限られた漢字(語構成要素)の組み合わせで多くの語が生み出されているということなのである。また、松下(2011, 2014)から分散度を計算すると、漢字は語に比べればジャンルの偏りが少なめで、一つの字がより多様なジャンルに用いられる傾向にある。

初級で漢字の基本と基礎的な高頻度漢字を学習した後、中級では漢字を増やすとともに、その組み合わせによる語の成り立ちを学ぶことが、語彙学習負担を減らすために極めて重要である。以下、具体的に数点指摘したい。

第一に、一つの字の音と訓が漢字のイメージを媒介に結合していることが重要である。例えば「ジョウショウ（上昇）」「ショウシン（昇進）」が高頻度語「のぼる」と漢字「昇」のイメージで結合しているということである。この結合がないと「ジョウショウ」「ショウシン」等の語意を別々に記憶することになり、学習項目数が増大して負担が増える。これは話しことばの語彙習得でも重要である（酒井, 2005; 松下, 2018）。例えば「上昇」の「上」には「場」「状」など多数の同音字があり、「昇」も「賞」「章」など多数の同音字がある。「上」「昇」が字形イメージを媒介に/ジョウ/, /ショウ/という読みとリンクしていれば、文脈の助けで漢字の特定と意味理解が促進され得る。音読みの二字漢語を和語の類義語で言い換えたり（上昇→のぼる）、和語の基本語から二字漢語を派生させたり（のぼる→上昇, 昇進）する練習は既習語を使えば効果的であろう（未習語が多いと逆効果になる可能性がある）。

第二に、語構成の知識（統語、比喻などの意味、音韻など）が重要である（Koda & Mitsugi, 2007 など）。例えば、2 字漢語の意味と単漢字の意味や結合パターンを考える活動は有効であろう。「入門（する）」の意味を確認した後で、「門」と「入（る）（ニュウ/はい（る））」の関係や「門に入る」と「入門する」の意味関係を考えたりする活動である。また、「高-」「-館」などの漢語系接辞の学習も重要である。

第三に、漢字の構成の知識を、既習漢字数が増えた段階で学び直すことも有効であろう。キヘンなどの部首（意符）はもちろん、「帳」「張」や「儀」「議」の共通部分が同じ音符を有するといった知識も有用であり、「紙」「組」「細」と「糸」の関係を考えるなど、部首と漢字の間に働く意味関係を考えることも有用であろう。

上述の諸点を学習者自身が体得するためか、1 語、1 字あたりの平均学習時間は中上級では初級や超上級より短い（松下ほか, 2021）。整理して短時間で体系的に取り上げる授業や教材があれば、非漢字系学習者にはさらに有効ではないだろうか。

3. 日本語語彙学習への学習者言語の影響

3.1 学習者の第一言語によって語種の相対難度はどう異なるか

松下ほか（2021）は、第1言語（L1）別に和語、漢語、外来語といった語種別のラッシュ項目難度を算出し、中国語 L1 には「相対的に」漢語は易しく外来語は難しい、韓国語 L1 とその他 L1 にとって外来語が易しく、日本語 L1 にとって和語が易しいという結果を出している。なお、松下・佐藤ほか（2020）は、中国の朝鮮族大学生は韓国語 L1 より中国語 L1 に近く、韓国語 L1 はむしろ非漢字圏 L1 に近いとしている。

3.2 漢語（や外来語）の学習への L1 の効果と、学習・教育のあり方の提案

松下・陳ほか(2020)が「日中対照漢字語データベース」を作成し、分析したところによれば、頻度上位2万語のうち、約35%がいわゆる同形漢語で、同形漢語の18% (同形漢語の5~6語に1語)が同形類義や同形異義といった要注意の語だと報告している。逆に言えば2万語のうち3割程度は日中同形同義の語である。

また、佐藤ほか(2020)は学術共通語彙テスト(JCAWT)の音声版と文字版を作成し(<http://www17408ui.sakura.ne.jp/tatsum/webtest.html#JCAWT>で受験可)、中国語L1と韓国語L1を対象に実施し、両グループとも文字版の得点のほうが高かったが、中国語L1のほうが文字版と音声版の差がより大きいことを報告している。その他のL1にこのテストを行えば、音声版の得点が高くなると予測される。

以上などから、中国語L1の日本語語彙学習について、以下の提案をしたい。

- ・基本義や共起成分が類似している語が多いので、文字と音声の同時処理による付随的語彙学習は効果的かもしれない。例えば、スクリプト付きの多聴、字幕付き映画の視聴、同一トピックの読解と聴解などが効果的だと考えられる。ふりがな付き多読も効果があるかもしれない。
- ・基本義や共起成分が異なっている場合は違いを意識的に克服する必要がある。漢字を見た瞬間に中国語知識は自動的に活性化するからである(ただし、習熟度が高まるにつれてL2知識はL1から独立していく(茅本, 2000など))。
- ・L1知識を活かして「語彙学習先行モジュール」(松下, 2002; 松下, 2005)を構想するのもよいと思われる。これは中国語L1に対して、語彙学習を先行させて、既習語彙を豊富な例文による文法学習などに生かすという提案である。

なお、英語知識の日本語の外来語学習への応用については検証が必要だが、現時点では、英語L1知識は活用できるが、英語L2知識のプラスは少ないと考えている。

4. まとめと今後の課題

漢字と語彙の知識の多様な側面の結合で語彙知識は安定する。非漢字系学習者には漢字を結節点にした音と訓の知識の結合や、統語・意味等に関わる語構成の意識的学習が効果的だと思われる。一方、中国語L1学習者には、文字と音声の同時処理による付随的語彙学習や両語の違いに関する意識的学習が効果的であろう。漢字語の音声インプットを教師も学習者もより意識して、音声と文字の語彙知識をスパイラルで高めていくようなカリキュラムについて、語彙学習先行など、文法学習との関係も含めた議論が必要である。例えば、カリキュラムの一部をモジュール化して、漢字系と非漢字系の学習者がそれぞれ選択できるカリキュラムを作ることも一案であろう。

日本語語彙の学習や教育のあり方の探求には、世界の応用言語学の先行研究は枠組みとして大いに参考になる。が、日本語の表記体系の複雑さを踏まえると、日本語教育研究者が、自分たちで考えねばならない問題も多い。例えば、ふりがなは漢字の読みの付随的語彙学習にどの程度効果的なのか、どのレベルの学習者にどの程度のふりがなつきテキストを与えるのが効果的か、など、まだまだ考えるべきことは多い。

参考文献

- 茅本百合子 (2000). 「日本語を学習する中国語母語話者の漢字の認知-上級者・超上級者の心内辞書における音韻情報処理」 『教育心理学研究』 48: 315-322.
- 酒井たか子 (2005). 「音声情報の関わる漢字能力とその測定の試み」 『筑波大学留学生センター日本語教育論集』 20: 45-56.
- 佐藤尚子・松下達彦・笹尾洋介・田島ますみ・橋本美香 (2020). 「学術共通語彙に関する音声知識と文字知識の違い —中国語および韓国語を第一言語とする日本語学習者に焦点を当てて—」 『第22回専門日本語教育学会研究討論会誌』 26-27. http://stje.kir.jp/download/22STJE_discussion.pdf
- 松下達彦 (2002). 「中国語を母語とする日本語学習者のための語彙学習先行モジュールの提案 ～第二言語習得理論, 言語認知, 対照分析, 語彙論の成果を踏まえて～」 《日語学习与研究》, 2002年 第1期: 50-54.
- 松下達彦 (2005). 「語彙学習先行モジュールの日中バイリンガル児童・生徒への応用 —母語の漢字知識を活かす—」 『母語・継承語・バイリンガル教育研究』 創刊号: 84-95.
- 松下達彦 (2011). 「日本語を読むための語彙データベース (VDRJ) Ver. 1.1」
<http://www17408ui.sakura.ne.jp/tatsum/database.html#vdrj> (2021年10月30日閲覧)
- 松下達彦 (2014). 「現代日本語文字データベース (CDJ) Version. 2.0」
<http://www17408ui.sakura.ne.jp/tatsum/database.html#cdj> (2021年10月30日閲覧)
- 松下達彦 (2017). 「日本語・教育・語彙 第14回 ことばの時事問題(1): なぜ『漢字ドリル』なのか」三省堂『Word-Wise Web』 <https://dictionary.sanseido-publ.co.jp/column/goi14> (2021年10月30日閲覧)
- 松下達彦 (2018). 「日本語語彙習得に関わる普遍性と個別性 —漢字をめぐる問題を中心に—」 『第四回学習者コーパス・ワークショップ&シンポジウム —第二言語習得における語彙の役割—』 発表資料, 学習者コーパスに基づく第二言語としての日本語の習得研究, 国立国語研究所
- 松下達彦・佐藤尚子・笹尾洋介・田島ますみ・橋本美香 (2020) 「学習者言語が日本語学術共通語彙の理解に与える影響 —中国語母語, 中朝バイリンガル, 韓国語母語, 非漢字圏の学習者を比較して—」 『専門日本語教育研究』 22: 25-32.
- 松下達彦・佐藤尚子・笹尾洋介・田島ますみ・橋本美香 (2021). 「日本語の語彙量と漢字力 —第一言語と学習期間の影響—」 『日本語教育』 178: 139-153.
- 松下達彦・陳夢夏・王雪竹・陳林柯 (2020). 「日中対照漢字語データベースの開発と応用」 『日本語教育』 177: 62-76.
- Koda, K. & Mitsugi, S. (2007). Word knowledge development in a second language. *Acquisition of Japanese as a Second Language* (『第二言語としての日本語の習得研究』), 10, 87-113.
- Matsushita, T. (2012). *In what order should learners learn Japanese vocabulary? A corpus-based approach* [Doctoral dissertation, Victoria University of Wellington].
<http://researcharchive.vuw.ac.nz/xmlui/handle/10063/4476> (2021年10月30日閲覧)