

語彙・漢字頻度表示器 J-LEX によるリライト教材作成
CREATING LEXICALLY CONTROLLED MATERIAL TEXTS BY J-LEX,
AN ONLINE LEXICAL ANALYZER OF JAPANESE TEXTS

松下達彦（東京大学）

Tatsuhiko Matsushita, The University of Tokyo

要旨：内容重視の日本語教育や、流暢さを上げるための速読、あるいは対象言語への接触を増やすための多読などが近年、注目を集めているが、これらの活動を円滑に学習活動に取り入れるためには生教材と同レベルの内容を保ちながら、語彙や文法のレベルを学習者のレベルに合わせて的確にコントロールすることが必要である。本稿では、特に語彙面でのコントロールの必要性に関する理論的な背景を説明した上で、オンライン語彙・漢字頻度プロファイラーJ-LEXを利用して、低頻度語彙・漢字などのリライトやふりがな付与による読解テキストの作成方法について説明する。

キーワード： 速読 多読 内容重視の日本語教育 リライト教材 J-LEX

1. はじめに

近年、内容重視の第二言語教育（CBI: Content-based Instruction）や速読、多読の実践が広がりを見せている。岡崎(1994,2002)は一連の著作で内容重視の日本語教育の重要性を主張し、その形態の一つとして多読を取り上げている（岡崎 1994）。また、欧米諸国でも言語教育における文化の重要性に対する認識が広まるにつれ、内容重視の日本語教育が繰り返し取り上げられている。例えば、2011年の18th Princeton Japanese Pedagogy Forum ではメインテーマとして内容重視の日本語教育が取り上げられ、例えばハドソン遠藤（2011）は短編小説の講読コース設立を勧めている。文学ばかりでなく、時事的な話題や専門分野など多様な内容で授業ができれば、より有意義な内容重視の授業ができるであろう。言語がコミュニケーションの手段の一つであることを考えれば、コミュニケーションの目的そのものを言語教育の目的とすることは自然であるし、批判的思考、創造性、自律性の獲得、あるいは地球的な視野や国際社会の一員としての知識や行動力の獲得といったより大きな教育目標に近づく意味でも、内容重視の活動が広がることは望ましい。

しかし、内容重視の日本語教育においては、言語の運用レベルを考慮に入れないわけにはいかない。話し合いなど、産出を中心にした活動であれば、運用レベルに基づいたクラス分けがなされている場合、言語運用レベルが自然に調節され、クラスメート同士が理解できないことは少ないであろうし、仮にそうであっても相互に尋ねて確認することができる。しかしながら、読解や聴解のような受容的な活動の場合、対象となるテキストの語彙や文法のレベルが高すぎると、内容重視の活動は成り立たない。多くの生素材は、そのままでは語彙や文法の負荷が高すぎて中級学習者には使えない。リライトは内容重視の言語教育、特に中級学習者のためには非常に重要である。

また、行政サービスにおける「やさしい日本語」（例えば横浜市「やさしい日本語での情報発信について」<http://www.city.yokohama.lg.jp/lang/ej/kijun.html>）においても、語彙レベルのコントロールがきわめて重要であることは言うまでもない。（筆者は2014年度より、科研費基盤研究A「やさしい日本語を用いた言語的少数者に対する言語保障の枠組み策定のための総合的研究」（代表者：庵功雄）に加わり、横浜市の広報課と協力して「やさしい日本語」による情報発信の語彙面での研究に参加している。）

2. リライトにおける語彙レベルコントロールの理論的な基礎

リライトで最も重要な理論的基礎はテキスト中の既知語量（カバー率）と読解レベルの関係である。英語では（研究によって異なるが）55%~70%の内容理解に、テキストの延べ語数に占める既知語のカバー率95%、自力で読解を楽しむには98%程度必要だと言われる（Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010 など）。日本語では、小森ほか(2004)は96%あたりに読解レベルの閾値（読解レベルが急激に上がるポイント）が存在する可能性を示唆し、Matsushita(2014)は既知語数12000語程度でカバー率93%程度に達し、約7割の理解度になると推計している。これらの知見を総合すると、例えば、語彙量約5000語の学生が、上位5000語のカバー率93%未満の文章を読むのは無理がある。

辞書を引いて読めればよいと考える人もいるかもしれない。例えば、オンラインでポップアップ辞書や語彙リストの自動作成機能を利用することは言語の学習を大いに助けるに違いない。一方で、多読のような活動においては、何回も立ち止まって意味を調べたり考えたりしたのでは読むという行為自体が成立しないと言われている（Day & Bamford, 2002 など）。だから graded readers の開発において、語彙レベルが厳しくコントロールされるのである。また、速読においては、既知語率は100%であるべきだと考えられている（Nation, 2007）。途中で止まって考えることは学習にはなっても流暢さを上げること（fluency development）にはならないのである。さらに、習得という観点から考えても、focus on form の考え方に従えば、形式と同時に意味の処理に意識が向いていることが必要であり、そのためには上述のように一定レベルの既知語によるテキストカバー率に達していることが必要条件である。すなわち、内容重視の第二言語教育や多読、速読においては、言語形式の学習のための精読などよりも高いレベルの既知語率が求められるのである。学習者の語彙レベルを適切な方法で測定し、同時にテキストの語彙レベルを適切にコントロールすることによって、最も効率的な内容重視の活動が実現できるのである。

このほか、読解に占める下位能力のうち語彙力が最も大きな要素であるということ（Bernhardt, 2005; Matsushita, 2012 など）や語彙がおおよそ頻度順に習得が進むといったことも前提にしている。

3. J-LEX のベースワード

ベースワードとは、語彙・漢字頻度プロファイラーに仕込まれている語彙・漢字の頻度リストのことである。J-LEX では松下（2011, 2013）に基づいて作成されたベースワードが仕込まれている。このベースワードの有効性は、さまざまなテストコーパスにおけるテキストカバー率が、例えば旧日本語能力検定試験の級別語彙リストよりも高いことなどで実証されている（Matsushita, 2012）。

4. J-LEX を利用したテキストのリライト

図1 J-LEX 利用手順①



以下、リライトの語彙面での調整に焦点を当て、オンライン語彙・漢字頻度プロファイラー J-LEX (菅長・松下 2013) を利用してリライト教材を作成する方法について説明する。(図1)

- 1) J-LEX <http://www17408ui.sakura.ne.jp/index.html> を開く。
- 2) 分析したいテキストをウィンドウに貼りつける。
- 3) ロップダウンリストから使用したいパスワード(語彙または漢字) を選ぶ

図2 J-LEX 利用手順②

新しい言語を学ぶ楽しさとは何だろうか。覚えた表現を使ってみて通じたときの楽しさ、そこから人間関係ができていく楽しさ、情報が増えて世界が広がっていく楽しさ、などいろいろあるだろう。最近ではドラマ、アニメ、マンガ、歌、ゲームなどをきっかけに日本語学習を始める人が増えている。そのようなことは別の学習の楽しみの一つに、「気になる単語」との出会いかがある。いままで英語や中国語を勉強してきたが、その過程で、母語の日本語では一言で表せないような単語はいつか出会い、そのことばの意味を特別に意識するようになっていく。

例えば、中国語には「感(カエン) (gan) という語がある。イントナレーションは高くまっすぐに響く第1声である。これは「まみがある」という意味である。(うっかつ) 肉や魚はもちろんだが、シタケやタケノコなどのまみもこれで表せる。甘い、辛い、塩辛い、酸っぱいなど、日本語と同じようなことばもある。それに加えて、このgan(かん)には料理を

新しい言語を学ぶ楽しさとは何だろうか。覚えた表現を使ってみて通じたときの楽しさ、そこから人間関係ができていく楽しさ、情報が広がっていく楽しさ、などいろいろあるだろう。最近ではドラマ、アニメ、マンガ、歌、ゲームなどをきっかけに日本語学習を始める人が増えている。

そのようなことばの学習の楽しみの一つに、「気になる単語」との出会いがある。いままで英語や中国語を勉強してきたが、その過程で、母語の日本語では一言で表せないような単語はいつか出会い、そのことばの意味を特別に意識するようになっていく。

例えば、中国語には感(カエン) (gan) という語がある。イントナレーションは高くまっすぐに響く第1声である。これは「まみがある」という意味である。(うっかつ) 肉や魚はもちろんだが、シタケやタケノコなどのまみもこれで表せる。甘い、辛い、塩辛い、酸っぱいなど、日本語と同じようなことばもある。それに加えて、このgan(かん)には料理を

表1 ジャンル/レベル別語数 (延べ)

特徴領域数 (0~4)	0~4 D	1 D	1 D	1 D	2 D	2 D	3 D / 4 D	2 D	1 D	1 D	0 D
ジャンル領域	全体	文芸	人文系	社会系	文系共通	文理混合系	学術共通	理系共通	理工系	生物・医学系	その他
※想定既知語彙	16	0	0	0	0	0	0	0	0	0	8
初級 (1-1285位)	743	17	4	0	2	2	36	0	0	0	699
中級前半 (1286-4000位)	150	13	4	3	18	7	70	0	1	0	47
中級後半 (4001-8000位)	31	0	2	0	0	4	0	0	0	0	25
上級前半 (8001-12000位)	11	0	0	0	1	0	2	0	0	0	8
上級後半 (12001-16000位)	1	0	0	0	0	0	0	0	0	0	1
超上級 (16001位以上)	7	0	1	0	1	0	0	0	0	0	5
リストに該当なし	101	0	0	0	0	0	0	0	0	0	0
合計	1052	30	11	3	22	13	108	0	1	0	793

- 4) 語彙レベル/漢字レベル等を数字で設定する。(例: [3000] 位以上を赤でハイライト)
- 5) 判定ボタンを押す。
- 6) 右側の結果ウィンドーを見て、適切なテキストカバー率が達成されていない場合、赤でハイライトされている語を減らすように左側のウィンドーのテキストを書き直し、再び判定ボタンをクリックする。(リライトの過程を残したい場合は、右側の結果ウィンドーのテキストをコピーし、MS-Wordなどにペーストする。色もそのまま保持される。その後、「変更履歴の記録」の機能を使用する)。

- 7) テキストが適切な語彙レベルに調節された後に、調整済みテキストを再びウィンドーに貼り付け、今度は漢字レベルを指定して判定する。指定レベルよりも頻度の低い漢字は、1) ふりがなをつける、2) かな表記にする、3) 他の語に置き換えるなどの対応をとる。

これらのほか、J-LEX は頻度レベルごとの延べ語数や異なり語数、累積テキストカバー率などをジャンル別に返す機能がある (図 2)。1000 語レベル別に語数、カバー率などを表で示す機能もあり、特に累積テキストカバー率を見れば、どのレベルの語をリライトすればよいか、どのレベルの語を学習のターゲットにすればよいか分かる。

J-LEX は、上述のような教材開発や各種サービスにおける「やさしい日本語」作成のための利用のほか、学習者等の作文の語彙・漢字レベルの判定や、試験や研究のためのテキストの語彙・漢字レベルのコントロールに使用することも可能である。

参考文献

- Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25, 133–150.
- Day, R. & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14(2), 137-141.
- Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Matsushita, T. (2012). *In What Order Should Learners Learn Japanese Vocabulary? A Corpus-based Approach*. Unpublished PhD Thesis. Victoria University of Wellington.
- Matsushita, T. (2014). How is the relationship between vocabulary knowledge and reading comprehension? A case of Japanese. *AILA World Congress 2014*.
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2–13.
- 岡崎 眸 (1994) 「内容重視の日本語教育 —大学の場合—」『東京外国語大学論集』 49, 227-244.
- 岡崎 眸 (2002) 「内容重視の日本語教育 —多言語多文化共生社会における日本語教育の視点から—」岡崎眸 (編) 科学研究費補助金研究成果報告書『内省モデルに基づく日本語教育実習理論の構築』322-339 【一部加筆修正】, URL: <http://www.dc.ocha.ac.jp/comparative-cultures/jle/Okazaki/naiyoo-juushi.html>, 最終確認日: 2015 年 2 月 15 日
- 小森和子・三國純子・近藤安月子 (2004) 「文章理解を促進する語彙知識の量的側面 —既知語率の閾値探索の試み—」『日本語教育』 125, 83–92.
- 菅長陽一・松下達彦 (2013) 「日本語テキスト語彙分析器 J-LEX」 URL: <http://www17408ui.sakura.ne.jp/index.html> 最終確認日: 2015 年 2 月 15 日
- ハドソン遠藤 (2011) 「CBI コースとしての短編講読講座」18th Princeton Japanese Pedagogy Forum, URL: http://www.princeton.edu/pjpf/past/18th-pjpf/05_EndoPJPF2011.pdf 最終確認日: 2015 年 2 月 15 日
- 松下達彦 (2011) 「日本語を読むための語彙データベース」 URL: <http://www.geocities.jp/tatsum2003/> 最終確認日: 2015 年 2 月 15 日
- 松下達彦 (2012) 「「日本語を読むための語彙量テスト」の開発」『2012 年日本語教育国際研究大会予稿集』第一分冊, 310.
- 松下達彦 (2013) 「現代日本語文字データベース」 URL: <http://www17408ui.sakura.ne.jp/tatsum/> 最終確認日: 2015 年 2 月 15 日