

日本語読解テキストのリライトの重要性とアプローチ

—語彙的要素を中心に—

松下 達彦

国際交流基金日本語国際センター・政策研究大学院大学

『日本言語文化研究会論集』第13号（2017年）

2017年10月 発行 抜刷

日本語読解テキストのリライトの重要性とアプローチ — 語彙的要素を中心に —

松下 達彦

要旨

近年、リライトの重要性が注目されている。本稿では、初めに、動機づけ、実用性および読解学習と語彙学習の観点からリライトの重要性を論じる。特に、1) 読解に占める語彙の重要性、2) 既知語率と読解レベルの関係、3) 精読・多読・速読において異なる既知語率が求められること等について先行研究を概観し、学習負担の調整、言語習得の可能性の拡大の観点からリライトの重要性を論じる。続いて、リライトに使えるウェブツールを紹介し、リライトの方針と方法を述べる。そのうえで、読解テキストの語彙学習可能性指標 LEPIX を提案し、リライト例および LEPIX の計算例を紹介する。最後に、残された課題として、テキストの語彙的難度の数値化の問題と、母語知識の利用による想定既知語をどうリライトの方針や方法に反映するかという問題について触れる。

〔キーワード〕 リライト、語彙学習、既知語率、語彙的難度、LEPIX

1. はじめに

近年、日本語教育においてリライトの重要性について論じられる機会が増えている。本稿でいうリライトとは「文法や語彙などのレベルをコントロールして、対象とする学習者が無理なく読めるように書き直すこと」である（本稿では「テキスト」を「文章」とほぼ同義で用いる）。リライトが注目される背景には、多読の実践（熊田・鈴木 2015；江田・飯島・野田 2005；高橋 2016；中野 2013；三上・原田 2011；Hitotsugi & Day 2004 など）、「やさしい日本語」（庵 2016；佐藤 2004；田中・美野 2016 など）への注目、テキストの特性から読みやすさを予測するリーダビリティ研究（柴崎・原 2010；柴崎 2014；李 2016 など）などがある。

本稿ではリライトの重要性について、動機づけの観点、実用性の観点、読解学習と語彙学習の観点から述べたうえで、具体的なリライトの方針や方法について、いくつかのツールの使い方を交えて述べる。そのうえで、残された課題について語彙的難度の数値化やツールに望まれる機能などについて、母語知識の利用の影響を含めて論じる。

2. リライトテキストの重要性

2.1 動機づけの観点

通常、選択外国語科目では中級、上級と進むにつれて学習者数が減る。日本語のカリキュ

ラムに即して考えると、中級、上級と上がっていくにつれて一番問題になるのは漢字や漢字語の学習だと思われる。日本語は文字学習の負担の大きい言語である。複数の文字種を混ぜ書きにする表記体系が複雑であることに加えて、漢字学習の負担が非常に大きいことが、上のレベルに進むことをあきらめる大きな理由の一つになっている¹。

漢字学習の困難は同時に読解の困難を意味する。2.3 で後述するが、一定の読解レベルに達するためには、一定レベルの既知語のテキストカバー率（知っている語の延べ語数がテキスト全体に占める割合）を必要とする。しかし、実際には、例えば初級で300程度の漢字を勉強しただけで読める生のテキストはほとんどなく、中級を終えて1,000字程度習得して、ようやく生のものである程度読める程度である。つまり、漢字を1,000字以上習得するまでは、生で読めるものがほとんどないのに、漢字の学習を延々と続けなければならないのである²。会話ならば各種のストラテジーを用いてインプットを調整し、わかる言葉で話してもらうようにできる。それに対して、通常読解では、頼りになるのは辞書だけであり、しかも頻繁に辞書を引く状況では読解という行為そのものが成立しないと言える。

ここにリライトの意義がある。まず、読解教科書のテキストを、未知語の量をコントロールして易から難へと無理なくレベルを上げられるように配列する工夫が必要である。しかし、これまで中上級の読解教科書で語彙レベルを測定、明示した教科書はあまり見たことがない。また、graded readers（段階的読み物）を使った多読も動機づけを保つ有効な方法である³。筆者が担当した多読クラスのコメントシートを見ると、初級であっても既習知識を使って自力で読めるということの喜びが見て取れる⁴。加えて、自分が学習した漢字が意味を持った文脈に現れることで、学習の効果を実感していることもわかる。もちろん、中級の読解教科書などにも多くのリライトテキストが使われているが、教科書だけではテキストに触れる量も質も圧倒的に限られていると言わざるを得ないし、教科書を超えて楽しんで読めるものが欲しいというのが学習者の本音ではないだろうか。

漢字学習を生日本語素材を読む準備として延々と行うのではなく、動機づけを保つため、漢字語学習に報いる報酬として、レベルに合った楽しめるリライトテキストが求められる。

2.2 実用性の観点

国際交流基金（2016）によると、2015年の時点で、世界には360万人以上の日本語学習者がおり、文化庁（2016）によると、2016年の時点で日本国内には20万人以上の学習者がいる。過去に学習をしていた人や、教育機関外の学習者で統計に表れない人を含めれば、簡単な日本語なら理解できるという人は少なくとも数千万人には及ぶと考えられる。日本語使用者とは難解な表現をすべて理解するような人ばかりをさすのではなく、さまざまな異なるレベルで理解する人を含むものである。そう考えると、理解しやすい日本語の使用域を拡大することは、社会的にコミュニケーション領域を拡大することに直結するのである。

庵 (2016) などで注目が高まっている「やさしい日本語」には多文化共生理念も反映されるが、実用的には例えば日本語初級レベルの住民にも伝わる自治体ウェブサイト構築という形で実現している (横浜市市民局広報課 2017)。公的サービスの多言語化はやさしい日本語化より早く実現しているが、住民には英語や中国語などの多言語サービスでカバーされない母語の人も多く、その場合に最もよく伝わる言語は「やさしい日本語」である。この考え方は災害時に必要な情報を必要な人に届ける方法として早くから提唱されていた (佐藤 2004)。

こうした人権ベースの公的サービスは言うまでもないが、街の中の各種サービスやマスメディアの情報をわかりやすい日本語で提供することにも大いに実用的な意味がある。電気店、雑貨の量販店、デパートなどの商業サービスや各種法人の公的サービスなどで多言語のチラシやパンフレットが作成されているが、「やさしい日本語」によるものもあれば、そのほうが助かる人も多いに違いない。NHK (日本放送協会) の NEWS WEB EASY はやさしい日本語によるニュースサイトであるが、中級レベル以下の日本語使用者にとっては貴重な情報源になっているはずである。英語では、Simple English Wikipedia や平易な英語で書かれた聖書など、限られた語彙で読めるものがたくさんある。日本語によって多様な情報やアイデアをより多くの人と共有するためには、適切に調整されたリライトテキストを使うことが有効であろう。

教育機関での日本語教育も、大学や日本語学校では、学習者が上級まで学習を続けることを前提にしたようなカリキュラムが多いが、実際には途中でやめる学習者のほうが多い。では、途中でやめた学習者にとって、日本語学習の意義は何であろうか。異なる言語に触れることの価値は、実用にとどまらず、認知的な発達や文化に関する発見にもつながるであろう。また、いわゆる旅行会話だけならば、初級を終えただけでも実用に足りるかもしれない。しかし、初級や中級のレベルでも成り立つコミュニケーションの空間を音声だけではなく、読み書きにおいても拡大する必要があるのではないだろうか。

文学、アニメやマンガにおいても、限られた文法、語彙で楽しめるものがあれば、学習だけのためでないという意味で実用的意義を持つといえる。それを実現しているのが NPO 法人多言語多読による「レベル別日本語多読ライブラリー (にほんご よむよむ文庫)」や、「さくら多読ラボ」⁵、Oxford Brooks 大学のグループが監修する Let's Read Japanese⁶ などの多読用の読み物シリーズである。エンターテインメントだけではなく、ニュースや社会的なトピックに関するエッセイなど、より多種多様な内容を、初級や中級レベルの日本語で読むことができれば、より豊かな内容の日本語学習ができるに違いない。

多文化共生のための行政サービスや住民の助け合いとしての「やさしい日本語」が必要とされるほか、初級、中級レベルの日本語使用者が多いという事実を踏まえ、商業サービス、マスメディアや各種の公的サービス、文学、エッセイなど、より幅広いジャンルでリライトテキストがあれば、日本語の実用性はより高まり、より豊かな日本語空間が広がるであろう。

2.3 読解学習と語彙学習の観点

ここまでは学習心理や社会的ツールの側面からリライトの意義を考えてきた。ここでは学習素材としてリライトテキストを、より応用言語学的な観点から考えてみる。

2.3.1 読解における語彙の重要性

以下、第二言語の読解力を構成する下位能力に関する代表的な研究を見てみる。

Bernhardt (2005)は、関連の論文を俯瞰したうえで、形態・統語的知識が読解力の約30%を占めるとし、特に語彙力の重要性を強調している。Bernhardtは全体を100%としたうえで下位能力の割合を論じている。Jeon & Yamashita (2014)は、いわゆるメタ分析の論文で、4,000以上の論文から基準を満たした59の論文を選んで再分析している。読解力と最も相関が高いのは文法力と語彙知識(既知語彙量)で、相関係数はそれぞれ.85(対象論文数13)と.79(対象論文数31)である。平均値は文法力のほうが上だが、語彙知識のほうが文法力よりも読解との相関が高いとする論文も多い。以下、聴解力、形態素の知識、符号化(decoding、音韻処理)、正書法(スペリング)知識、音韻認識、第一言語の読解力、作業記憶、メタ認知と続く。ここでは全体を100とするのではなく、読解力との関係を相関で示している。

日本語を対象とした研究はどうであろうか。Koda (1989)は、言語知識や処理スキルの各種の側面と読解力の相関を報告しているが、読解力と最も相関が高いのは語彙知識で $r=.74$ ($p<.001$)としており、相関の二乗によって分散の55%が語彙知識で説明される。小森・三国・近藤(2004a)では、相関は全体で $r=.70$ ($p<.01$)だが、回帰分析を行って、既知語率が読解力の47%を説明できるとしている。野口(2008)は2005年の日本語能力試験の分析であるが、文字・語彙と読解の得点の相関(r)について、1級で.66、2級で.64、3級で.78、4級で.80と報告している。レベルが初級に近づくほど語彙力が読解力に占める割合は高い。最も低い2級レベルの相関を二乗しても分散の41%を説明できることになる。これらの研究を総合すると、少なく見ても読解力の4割程度は語彙力で説明できるといえるであろう。

2.3.2 既知語率と読解レベルの関係

第二言語学習者を対象とした、テキスト中の既知語率と読解レベルの関係を見てみよう。英語等を対象とした研究では、一定の読解レベルを保証する既知語率の閾値(threshold level)が存在するかどうか、存在するならばどの程度なのかが論争になっている(Hu & Nation 2000; Laufer & Ravenhorst-Kalovski 2010; Schmitt, Jiang, & Grabe 2011など)。Schmitt他(2011)は、明確な閾値は存在せず、既知語率と読解レベルの関係は直線的だとしている。しかし、上述のその他の研究では、適切な読解レベル(adequate comprehension)に達するための閾値もしくは必要な語彙サイズが存在すると主張されており、概ね95%から98%の既知語率が想定されている。例えば、Laufer & Ravenhorst-Kalovski (2010)は(い

ずれも固有名詞を含めて) 最低 95%、最適レベルとして 98%を主張している。小森・三國・近藤 (2004b) は日本語を対象に研究し、96%あたりに閾値がありそうだと結論づけている。

2.3.3 目的に応じた既知語率や未知語の頻度の調整

ここでは閾値の有無は議論しないが、いずれにしても、既知語・未知語のレベルをコントロールすることは、適切な読解テキストの選択にとって、以下の点で非常に重要である。

第一に、目的によって必要とされる既知語率が異なるということである。Nation & Deweerdt (2001)はうまくリライトしたテキストが meaning-focused input (意味に焦点を当てたインプット) と fluency development (流暢さを上げる練習) に欠かせないと主張している。

精読では、ある程度理解できるが少し努力が必要というレベルにコントロールするべきである。通常は 95% (未知語は 20 語に 1 語) 前後の既知語率があれば、ある程度内容が理解でき、適度に未知語を含むので、辞書等の助けを借りて読む精読教材として使えるであろう。

多読など、他者や辞書の助けを借りない読解に十分な既知語率は、一般には 98%程度 (未知語は 50 語に 1 語) といわれる。近年、CBI や CLIL など、内容重視の言語教育が提唱されることが増えている (佐藤・柴田 2016 など)。筆者も批判的思考や創造性を育てる言語教育を心がけ、そのような素材を探して各種タスクを考えるが、その大前提は、使用される言語が理解できるということである。内容重視の授業を効率的に行うには、通常精読に必要な既知語率よりも高い既知語率が求められる。

速読の授業では、原則として 100%の既知語率が求められる (Nation 2007)。流暢さを高める練習では、すべて理解できるテキストをいかに短時間で理解できるかが問題だからである。筆者が授業で速読の練習をしたところでは、400 字程度の初級レベルのテキストを理解するのに、初級学習者で 3~4 分、中級で 2~3 分、上級では 1~2 分とだんだん短くなる。読解のレベルだけではなく、言語処理の速度を上げることも重要である。

第二に、学習負担を調節する必要性である。一般的に語彙学習は学習者任せにされることが多い。筆者は、留学中に 1 週間にいくつの英単語を覚えられるか、試したことがある。20 語ぐらいは問題なく覚えられた。1 週間 20 語なら、年間で 1,000 語程度となり、通常の留学生ならばこのぐらいは期待できる。1 週間に 50 語以上詰め込むと、数週間でかなりの単語は忘れていく。未知語の異なり語数が多すぎると、学習負担は増す。延べ語数は読解レベルと関わるが、異なり語数は学習負担に直結するのである。学習者の負担が十分に考えられていない教育現場をしばしば見かけるが、語彙学習の苦しさは自分が学習者になればわかる。教師はもう少し真剣に学習負担を適度に調節することを考えるべきである。

第三に、意味推測と言語習得の可能性を高めることである。近年、文脈からの未知語の意味推測や、読解などの活動に伴う付随的語彙学習 (incidental vocabulary learning) の研

究が進んでいる。既知語率が低ければ、当然、意味推測はうまくいかなくなる。また、既知語率が同じでも、未知語の出現頻度によって、習得レベルは異なる。付随的語彙学習の状況では、未知語の習得に必要な出現頻度は少なくとも2~6回 (Rott 1999)、多ければ20回以上 (Waring & Takaki 2003) とされる。語彙学習の効果を高めるためには、ただ未知語を登場させればよいのではなく、できるだけ繰り返し登場させるべきなのである。

これまで、このような点に配慮した中上級教科書はあまり見たことがない。通常のテキストではレベルが上がるにつれて、未知語の出現回数が少なくなり、キーワード以外は多くが one-timer と呼ばれる出現回数1の語となる。これが中上級の学習者を苦しめる。日本語能力試験 N1 の出題範囲は15,000語程度とされているが、15,000位レベルでは、100万語あたりの出現頻度は3回にも満たない(表1)。手元の文庫本や新書の字数と語数を計算してみると、概ね1冊あたり12万~18万字で、7万~11万語(形態素)程度である。平均で9万語程度だとすると、N1レベルの語は、自然な状況では本を3冊か4冊読んでやっと1回だけ出てくるだけである。それを数千語も覚えなければならないのである。したがって、少しでも習得しやすくするため、学習目標レベルの語の出現回数を可能な限り増やすリライトが望ましい。

表1 『日本語を読むための語彙データベース』(留学生用ランク)の1,000語レベルごとの平均語彙頻度

レベル	語彙順位	100万語 あたり 平均頻度	語彙レベル (Kは1000語ごと のまとまりを表す)	100万語あたり 平均頻度
初級	1-1,291	548.5	01K	694.6
			02K	102.0
中級	1,292-5,000	46.5	03K	40.6
			04K	23.1
			05K	15.3
			06K	11.5
			07K	9.1
中上級	5,001-10,000	7.9	08K	7.4
			09K	6.2
			10K	5.2
			11K	4.4
			12K	4.0
上級	10,001-15,000	3.5	13K	3.4
			14K	2.9
			15K	2.7
			16K	2.3
			17K	2.1
超上級	15,001-20,000	1.9	18K	1.9
			19K	1.7
			20K	1.5

(出典: Matsushita 2012, p. 298 の表を日本語訳)

ここまで語彙に着目してきたが、文法習得にとっても既知語率の調整は重要である。いわゆる focus on form (Long 1991) の考え方に従えば、形式と同時に意味処理に意識が向いていることが必要である。十分に意味のつかめないテキストでは、意味処理に意識を向けることは難しい。理論的には未知語の意味を推測できるレベルのテキストで初めて形式と意味を同時に処理できると思われる。95%程度の既知語率は文法習得にとっても必要であろう。

3. テキストのリライトの方法

3.1 テキストのレベルをチェックするツール

テキストのリライトのためには、まずテキストの語彙や文法のレベルを把握しなければならない。ここでは、そのためのウェブツールをいくつか紹介する。

語彙レベル (のみ) のチェックのツール

1) Reading Tutor (Kawamura, Kitamura & Hobara 1997)

<<http://language.tiu.ac.jp/>>

旧日本語能力試験出題範囲に基づく語彙レベル・漢字レベルのチェックなどができる。辞書機能もある。ただ、旧日本語能力試験出題範囲の語彙リストがやや古く、初級語彙に「カセット」「マッチ」などの今はあまり使われないような語が数十語あったり、逆に現在では高頻度語彙になっている外来語が不足していたりする。また、レベルの区切りが四つしかないため、重要度のレベルを細かく判断することは難しい。

2) J-LEX (菅長・松下 2013)

<<http://www17408ui.sakura.ne.jp/tatsum/webtools.html#jlex>>

リライトのためのウィンドーがある。リライトレベルを頻度順位の1の位まで設定でき、簡便に使える。語彙頻度プロファイル(テキスト中の語彙の出現回数を頻度レベル別に示したもの)を1,000語ごとの区切りで表示する。漢字レベルもチェックできる。

文法レベル (のみ) のチェックのツール

3) 学習項目解析システム (筑波大学留学生センター)

<<http://lias.intersc.tsukuba.ac.jp/checker/>>

テキストを投入すると、文法項目を抽出してレベル判定を含む各種情報が出てくる。

4) はごろも (2012-2015「日本語教育のためのコーパスに基づく文法項目データベース構築と検索システムの公開」科研グループ) <<http://jreadability.net/hagoromo>>

機能語(助詞・助動詞など)を入れると、レベル判定を含む多彩な情報が出てくる。

5) あすなる (仁科喜久子・奥村学・傅亮・阿辺川武)

<<https://hinoki-project.org/asunaro/index-j.php>>

構造の解析、係り受けの表示ができる。レベルを数字で表示する機能はないのでリライ

ト向きではないかもしれない。

語彙と文法のレベルや読みやすさのレベルを総合的にチェックするツール

- 6) やさ日チェッカー文章診断版 (一般向け) (岩田・森・松下 2015)

<<http://www4414uj.sakura.ne.jp/Yasanichil/nsindan/>>

役所の文書をチェックするために作成されているが、それ以外の目的にも使用可能である。語彙については、初級、中級前半、中級後半、上級の4段階が色分け表示される。

- 7) jreadability (李在鎬・長谷部陽一郎・久保圭) <<http://jreadability.net/>>

文章全体の難易度の評価がある。語彙レベルのチェックもあるが、語彙レベルは複数の日本語教員による学習レベルの主観判定である。筆者がいくつかの単語を入れた結果を見る限りでは、頻度レベルとは異なるレベルの判定をすることが少なくない。

- 8) 日本語リーダビリティ測定 (柴崎秀子・原信一郎)

<<http://readability.nagaokaut.ac.jp/readability>>

テキストの難易度について、小中学校の学年レベルの判定ができる。ただし、800字以上の文章でなければ正しい判定は難しいようである。

3.2 リライトの方針・方法、および読解テキストの語彙学習可能性指標 (LEPIX) ⁷

3.2.1 リライトの方針と、その前提となる仮定

リライトは勘に頼ってもできるが、応用言語学的な知見に従えば、より効果的に行えるはずである。Ghadirian (2002)は、TextLadder というソフトウェアを開発し、出現回数の少ない語彙を減らすコントロールなど、語彙学習に適切なテキストのリライト支援を試みている。Cobb (2007)も WebVP という語彙頻度プロファイラーを開発し、頻度レベルをテキスト上で色分け表示する等の機能を搭載して頻度レベルに応じたリライト支援を行っている。

本研究では、これらの応用言語学的な知見も踏まえながら、以下の四つの方針を立てる。

- A) 読み手に既知と仮定される語の割合 (想定既知語率) を一定のレベル以上にする。
- B) 未知語と仮定される語が一定回数以上に出現するようにする。
- C) より多くの異なる学習目標語が残るようにする。
- D) 学習目標語がより高い割合で出現するようにする。

以下、この4点の裏付けとなる前提について説明する。

- A) 読み手に既知と仮定される語の割合 (想定既知語率) を一定のレベル以上にする。

上述のように、目的に応じて適切な既知語率を確保しなければ、精読も多読も速読も難しい。Laufer & Ravenhorst-Kalovski (2010) など、既知語率と読解レベルの関係を調べた研究が裏付けとなる。実際にどの程度が既知なのか知るには、例えば、文章の総語数をあらかじめチェックしておき、教室で学習者に未知語に線を引くようなタスクを課して数えてもらえばよい。その文書の既知語率は簡単に計算できるであろう。

B) 未知語と仮定される語が一定回数以上に出現するようにする。

「未知語と仮定される語」とは、例えば、既知語率を95%と考えた場合、残りの5%に相当する語のことである。「一定回数」については、テキストの長さにもよるが、普通は2回と考えればよいと思われる。通常の1回の授業で終えられる程度の長さのテキストでは、3回以上登場する語は非常に限られるからである。

2.3.3で触れたが、付随的語彙学習では、最低でも2~6回は接触しないと、未知語の習得は難しい (Rott 1999 など)。言語教育の場では覚えようという意識が働いていることも多いので付随的語彙学習とは言えないことも多いが、それでも、出現回数1の語の習得は大変である。読解の前後の聴解、ディスカッション、作文等で同じ語に出会えば習得の可能性は高まるが、それでもテキスト中の頻度は少しでも多いほうがよい。

C) より多くの異なる学習目標語が残るようにする。

これは、一見、2.3.3で述べた「未知語の異なり語数が多すぎると、学習負担は増す」ということと矛盾するように見える。しかし、それはone-timer (出現回数1の語)が多すぎるからである。実際には1)と2)で述べた方針に従って、一定の頻度レベルを超えていて2回以上出現する語のみを学習対象語と考えた場合、多くの異なり語を残すことは意外に難しい。1)と2)の方針を踏まえたうえで考えれば、学習目標語の異なり語数は多いほうがよい。そのほうがより多くの語を習得できるからである。

D) 学習目標語がより高い割合で出現するようにする。

密度は効率に関わる。いくら多くの学習目標語が出てきても、それが長いテキストを読んで初めて達成されるのでは効率的でない。限られた長さのテキスト中に学習目標語が豊富かつ頻繁に出てくるテキストが語彙学習のリソースとして優れている。

これらの方針を更にわかりやすくまとめると、テキストの語彙や文法を学習者のレベルに合うよう書き換えたりして、残された低頻度語彙のうち出現回数1の語を、①2回以上出現するよう書き換え、②より高頻度の語に書き換え、または、③削除、ということになる。

以上を踏まえて、3.1で紹介した「学習項目解析システム」と「J-LEX」を使ったリライト方法をまとめると以下のようなになる。

まず、「学習項目解析システム」で文法項目をチェックして、リライトする。

1. リライトしたいテキストを貼りつけ、難度の高い文法項目を書き換え、レベルに合った項目だけが残るまで繰り返してチェックする。一般的には長い文を少し短く分けるようなリライトが必要になることが多い。

次に、「J-LEX」で語彙のチェックをしてリライトする。

2. 分析したいテキストをウィンドーに貼りつける
3. ドロップダウンリストから使用したいベースワード (語彙または漢字) を選ぶ
4. 語彙レベル/漢字レベルを数字で設定する (例: [3, 000] 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)

高頻度語に入れ替える simplification (例:「語彙」→「語の集まり」)だけでなく、「語彙、すなわち語の集まり」というように、関係節や挿入句で語積的な表現を文章に混ぜ込む elaboration も使ってよい。

また、以下のような語については、低頻度語であったとしても既知語と想定して、既知語率に算入することを検討する。

- ・繰り返し使われるキーワード
- ・書きことばでは低頻度でも、話しことばでは高頻度でなじみがあると思われる語
- ・外来語や漢語で、学習者母語などの知識から正しい意味が理解できる語
例)「達観」「墓穴」(中国語)、「カウンセラー」「ゴシップ」(英語)
- ・高頻度の語構成要素の組み合わせで意味が容易にわかる複合語・派生語(意味的な透明度が高く、構成要素が高頻度の語)
 - ・単漢字の組み合わせで意味が容易にわかる漢字語 例)「両眼」「雨季」
 - ・和語の複合語 例)「突き倒す」「盗み出す」のような複合動詞

テキストが適切な文法と語彙レベルのレベルに調節されたら、最後に漢字のチェックをしてリライトする。

7. 調整済みテキストを再びウィンドーに貼り付け、今度はJ-LEXで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、

- a) ふりがなをつける、b) かな表記にする、c) 他の語に置き換えるなどの対応をとる。

なお、学習目標語のリストを作りたい場合、例えば頻度順に語彙を並べた場合の95～98%あたりを主な対象だと考えてリスト化し、98%以上の低頻度語彙は、できるだけ削除してしまうほうがよいケースが多いように思われる。

3.2.2 読解テキストの語彙学習可能性指標 (LEPIX)

3.2.1で述べたA)～D)は、実は数値化して考えることができる方針ばかりである。そこで、語彙学習に適した読解テキストかどうかを測定する指標値を考えてみた。

まず、(J-LEX等を使用して)一定の既知語率レベル(例えば95%)を想定し、それを超えるレベルの語のうち、2回以上登場する語を学習目標語と考え、その異なり語数(T)と

延べ語数 (W)、さらにはテキスト全体の延べ語数 (すなわちテキスト長) を数える (N)。

3.2.1 で述べた方針の D) で述べた学習対象語の密度は W/N である。C) で述べたように異なり語は多いほうがよいので、単純に T を乗じることで $W \cdot T/N$ という式ができるが、数字が小さくなりすぎるとわかりにくいので、単純に 100 を乗じると以下の式が得られる。

$$\text{LEPIX} = (T \cdot W \cdot 100) / N$$

これが「読解テキストの語彙学習可能性指標」LEPIX (Lexical Learning Possibility Index for a Reading Text) である。

LEPIX の使用にあたって、一つだけ注意しなければならないのは、長さが著しく異なるテキストの LEPIX 値を直接比較できないということである。テキストの延べ語数と異なり語数の関係が直線的に比例しないからである。これは幼児の語彙発達の研究者などの間で Type/Token Ratio (TTR) の計算法をめぐって多くの修正案が出ていること (Richards & Malvern 1997) に関係しているが、興味深いのは、TTR が幼児の語彙の豊かさを示す指標として考えられているのに対し、LEPIX はその逆だということである。すなわち、語彙が豊かすぎるテキストは、語彙学習にとっては不利なのである。LEPIX 値の高いテキストは洗練されていないかもしれないが、わかりやすく、語彙習得に結びつきやすいのである。

3.3 リライトおよび LEPIX 計算の例

例えば、図 1 のようなテキストを、中級後半レベルの学習者を対象に、上位 2,000 語のカバー率を 95% にすることにして、図 2 のように書き換えたとする。図 1 の網かけで示された部分は図 2 ではリライトされている。下線・太字で示された「コミュニティ」「囲う」「緊張」の 3 語は、リライト前は 1 回しか出現していないが、リライト後に 2 回出現するようにすることで学習目標語になっている。図 2 の「陰」(囲み文字) という語は、リライトの過程で、新たに学習目標語として追加された語である。「リビング」「たまる」(ゴシック体) の 2 語は、書き換えられてはいないが、全体のテキストレベルが下がったことで頻度上位 95% から外れ、結果的にこのレベルの学生にとって学習目標語となった語である。リライトの結果、削除された語は「キャンパス」「俗」「しっくり」「妨害」「空襲」などの中級後半以上の語を中心に 20 語ほどに及ぶ。

図 2 の文章でも易しく見えないかもしれないが、既知語数 2,000 語程度の中級前半学習者を想定すれば、読みやすさと語彙学習の可能性は図 1 より格段に上がっているはずである。

図 1、図 2 で示したテキストを例に LEPIX を計算してみると、リライト前は 95% に達するのに 4K レベルまでの語彙が必要で、LEPIX 5.8 だったものが、リライト後には、2K レベルで 95% に達するようになり、LEPIX は 36.5 になっている (表 2)。

テキストの長さにもよるが、1,000~2,000 字のテキストの場合、うまくリライトされたテキストであれば、95% 点を基準にした LEPIX 値は概ね 20~40 ぐらいになるようである。

学生の頃、建築学科の友人にキャンパス内で一番落ち着く場所はどこかと聞かれたことがある。この質問を折に触れて思い出す。人間には明確な目的もなく何となく集まれる場所が欠かせない。俗に言う「たまり場」である。

例えば、家族はなんとなくリビングに「たまる」ことが多い。自分の部屋にこもるのも悪くないが、一方で、なぜかリビングの方がじっくりくるときもある。手を伸ばせばお茶やお菓子があつたり、新聞や雑誌が転がっていたりして雑然としているが、それも悪くない。妨害されない限りは誰かがいても気にならない。

このような空間は、家族だけでなく、学校や会社や地域コミュニティでも必要ではないだろうか。私の故郷の名古屋は喫茶店が多いことで有名だが、これは第二次世界大戦の時に街の中心部が空襲で焼けてしまったからではないだろうか。人間は何もないところでは落ち着いて話ができないものだ。コミュニケーションには、木陰のような、囲われたスペースが必要にちがいない。なんとなくたまる場所として喫茶店は機能したのである。

理由がなくても集まれる、雑然とした場所で、話しても話さなくてもいい誰かがいる。それが「たまり場」の条件である。「誰か」は家族や友達かもしれないし、知らない人かもしれない。いずれにしても、不必要な緊張を強いられることのない相手にちがいない。学校にも、会社にも、近所にも、家族にも、「たまり場」は大切だが、もしかしたら、国と国、文化と文化の間にもそういう場が必要なかもしれない。

図1 リライト前のテキストの例（筆者作成）

学生の頃、ある友だちに大学の中で一番落ち着く場所はどこかと聞かれたことがある。この質問を今でもときどき思い出す。人間にははっきりした目的もなく何となく集まれる場所が必要である。よく言う「たまり場」である。

例えば、家族はなんとなくリビングに「たまる」ことが多い。自分の部屋で過ごすのも悪くないが、一方で、なぜかリビングの方が落ち着くときもある。まわりにお茶やお菓子があつたり、新聞や雑誌が置かれていたりして雑然としているが、それも悪くない。じゃまされない限りは誰かがいても気にならないし、緊張することもない。

このような空間は、家族だけでなく、学校や会社や地域コミュニティでも必要ではないだろうか。私の生まれた街は喫茶店が多いことで有名だが、これは戦争の時に街の中心部が焼けてしまったからではないだろうか。人間は何もないところでは落ち着いて話ができないものである。コミュニケーションには、木の陰のような、囲われた空間が必要にちがいない。なんとなくたまる場所として喫茶店は機能したのである。

理由がなくても集まれる、雑然とした、陰になった囲われた空間で、話しても話さなくてもいいような、じゃまにならない誰かがいる。それが「たまり場」の条件である。「誰か」は家族や友達かもしれないし、知らない人かもしれない。どちらにしても、不必要に緊張する必要のない相手にちがいない。学校にも、会社にも、地域コミュニティにも、家族にも、「たまり場」は大切だが、もしかしたら、国と国、文化と文化の間にもそういう場が必要なかもしれない。

図2 リライト後のテキストの例（筆者作成）

表2 LEPIX 計算の例

項目	リライト 前	リライト 後
テキストの長さ (= 延べ語数) (N)	361	373
異なり語彙素数	137	123
テキストカバー率95%を超えるレベルの延べ語数	18	18
テキストカバー率95%を超えるレベルの異なり語彙素数	14	9
テキストカバー率95%点語彙レベル = テキスト語彙レベル (LLT95)	4K	2K
95%テキストカバー率を超える学習目標語の最低生起数	2	2
テキストカバー率95%を超える学習目標語の延べ語数 (W95)	7	17
テキストカバー率95%を超える学習目標語の異なり語彙素数 (T95)	3	8
学習目標語密度 (%) (W95*100/N)	1.9	4.6
異なり学習目標語彙素生起回数 (W95/T95)	2.3	2.1
読解テキストの語彙学習可能性指標 (95%点基準) [LEPIX95] = (T95*W95*100)/N	5.8	36.5

3.4 リライト方針およびツールに関する残された課題

3.4.1 テキストの語彙的難度の数値化

これまでリーダビリティ研究では、テキストの難度の数値化が試みられてきたが、語彙的難度のみの数値化は十分に探求されていない。最も簡便な指標は低頻度語彙を含む割合や、一定の既知語率に到達するのに必要な語数であるが、これらの方法はある基準で区切ることでしか語彙的難度を示せないため、区切り内の語彙頻度の分布を表現しない。どこで区切るかによって語彙的難度が変わってくるのである。語彙頻度プロファイルを眺めれば頻度分布の傾向は直観的に理解できるが、その直観をなんとか数値で表現できないだろうか。例えば、あるテキスト中のすべての語について、大規模コーパスにおける頻度の対数の平均値はどうであろうか。あるいは、計算の過程で低頻度語彙や高頻度語彙に何らかの重みづけをするほうがより応用的に意味の数値になるかもしれない。そのような数値の計算をツールに組み込むことができれば、段階的読み物 (graded readers) の開発なども容易になるであろう。

3.4.2 母語知識の利用による既知語をどうリライトやテキストの語彙的難度に反映するか

よく知られているように、中国語系学習者は母語の漢字知識を利用してテキストをある程度理解できる。英語系学習者もカタカナの音韻符号化さえ正しくできれば、かなりの割合で母語知識が使える。既知語率計算では固有名詞を初めから算入することが多いが、これは多くの固有名詞が事前の学習を要求しない (Nation & Webb 2011: 166) からで、同じことが借用語についても言える⁸。借用語の原語の知識は、概ね初めから使える第二言語の語彙知識である。これをテキストの語彙的難度に反映させるには、漢字語や外来語のデータベースを作り、日本語の意味とのずれを数値化する必要がある。それができれば、あるテキストの語彙

難度が、中国語系学習者とそうでない人にとってどの程度異なるかを数値や画面イメージで示せるであろう。中国語系学習者向けのリライトとそうでないリライトを分けることができるはずである。そのようなデータベースを組み込んだ語彙頻度プロファイラーが望まれる。

4. おわりに

リライトの重要性は、動機づけ、実用性、語彙学習や読解学習の観点から説明される。日本語を使う空間が世界中に広がる一方、さまざまなレベルの日本語を使う人が増えている。そのため、リライトにはより科学的な知見とそれに基づくツールの開発が求められている。

謝辞

本稿の執筆にあたり、資料検索を劉瑞利さん（お茶の水女子大学大学院生）にお手伝いいただきました。田畑サンドーム光恵先生（Massey University）には草稿に貴重なご意見をいただきました。英文要旨についてはSky Marsen 先生（Flinders University）にもご意見をいただきました。感謝申し上げます（ただし、不備、誤り等があれば筆者の責任です）。

注

- ¹ 上のレベルまで続ける学生の割合を retention rate（継続率）と呼び、この割合が予算やスタッフの人数に反映されるとも聞く（例えば Martín, Jansen & Beckmann 2016）。
- ² Norton Peirce（1995）が学習者のアイデンティティを investment（投資）という語で概念化して以来、学習の継続や放棄がこの語で説明されることが増えた。
- ³ Yamashita（2013）や Banno & Kuroe（2016）に詳しい。
- ⁴ 松井・三上・金山（2012）など、多くの多読実践研究が学習者の喜びの声を伝えている。
- ⁵ <http://jgrpg-sakura.shobi-u.ac.jp/>（2017年8月19日参照）。JGRPG（Japanese Graded Reader Project Group）が運営している。
- ⁶ <https://www.brookes.ac.uk/lets-read-japanese/>（2017年8月19日参照）。
- ⁷ 3.2の内容は筆者のPhD論文（Matsushita 2012）の第8章に基づいている。
- ⁸ 欧米言語の研究では同根語効果 cognate effect と呼ぶことが多い。これは英語とラテン語の関係のように、同系統の言語間の影響を指すためであろう。

参考文献

- (1) 庵功雄（2016）『やさしい日本語—多文化共生社会へ』岩波書店
- (2) 岩田一成・森篤嗣・松下達彦（2015）『「やさしにちチェッカー」—公的文書の難易度を指標化するシステムの開発—』『2015年度 日本語教育学会秋季大会 予稿集』日本語教育学会、407-408.

- (3) 熊田道子・鈴木美加 (2015) 「日本語教育における Extensive Reading(多読)の実践」『東京外国語大学留学生日本語教育センター論集』41、229-243.
- (4) 国際交流基金 (2016) 「2015 年度海外日本語教育機関調査結果 (速報値)」 <<https://www.jpf.go.jp/j/about/press/2016/dl/2016-057-2.pdf>> 2017年9月17日参照
- (5) 小森和子・三國純子・近藤安月子(2004a) 「文章理解を促進する既知語率に関する一考察」『2004年日本語教育国際研究大会 発表1』日本語教育学会、167-172.
- (6) 小森和子・三國純子・近藤安月子(2004b) 「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.
- (7) 江田すみれ・飯島ひとみ・野田佳恵 (2005) 「中・上級の学習者に対する短編小説を使った多読授業の実践」『日本語教育』126、74-83.
- (8) 佐藤和之 (2004) 「災害時の言語表現を考える—やさしい日本語・言語研究者たちの災害研究」『日本語学』23(10)、34-45.
- (9) 佐藤雅彦・柴田健 (2016) 「CLIL を用いた上級日本語教育の試み—SFA をビジネス書で学ぶ総合クラスの事例—」『秋田大学教育文化学部教育実践研究紀要』第38号、221-232.
- (10) 柴崎秀子・原信一郎 (2010) 「12 学年を難易尺度とする日本語リーダビリティ判定式」『計量国語学』27(6)、215-232.
- (11) 柴崎秀子 (2014) 「リーダビリティ研究と「やさしい日本語」」『日本語教育』158、49-65.
- (12) 高橋亘 (2016) 「授業外日本語多読活動「日本語多読セッション」報告—3学期間にわたる活動実践を通して—」『東京外国語大学留学生日本語教育センター論集』42、185-196.
- (13) 田中英輝・美野秀弥 (2016) 「ニュースのためのやさしい日本語とその外国人日本語学習者への効果」『情報処理学会論文誌』57(10)、2284-2297.
- (14) 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源—形態素解析用電子化辞書の開発とその応用—」『日本語科学』22、101-123.
- (15) 中野てい子 (2013) 「電子教材を活用したブレンド型多読授業の導入」『尚美学園大学総合政策論集』17、137-144.
- (16) 日本放送協会 「NEWS WEB EASY」 <<http://www3.nhk.or.jp/news/easy/>> 2017年9月17日参照
- (17) 野口裕之 (2008) 「試験結果の分析」国際交流基金・日本国際教育支援協会編『平成17年度日本語能力試験 分析評価に関する報告書』凡人社、45-111.

- (18) 文化庁文化語部国語課 (2016) 「平成 28 年度 国内の日本語教育の概要」 <http://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/nihongokyoiku_jittai/h28/pdf/h28_zenbun.pdf> 2017 年 9 月 17 日参照
- (19) 松井咲子・三上京子・金山泰子 (2012) 「初級・中級日本語コースにおける多読授業の実践報告」『ICU 日本語教育研究』9、国際基督教大学日本語教育研究センター、47-59.
- (20) 三上京子・原田照子 (2011) 「多読による付随的語彙学習の可能性を探る－日本語版グレイディッド・リーダーを用いた多読の実践と語彙テストの結果から－」『国際交流基金 日本語教育紀要』7、7-23.
- (21) 横浜市市民局広報課 (2017) 「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>> 2017 年 8 月 14 日参照
- (22) 李在鎬 (2016) 「日本語教育のための文章難易度に関する研究」『早稲田日本語教育学』21、1-16.
- (23) Banno, E., & R. Kuroe (2016) Effects of extensive reading on Japanese language learning. In M. Gobert (Ed.) *Proceedings of the 3rd World Congress on Extensive Reading*.
- (24) Bernhardt, E. (2005) Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25, 133-150.
- (25) Cobb, T. (2007) Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, 11(3), 38-63.
- (26) Ghadirian, S. (2002) Providing controlled exposure to target vocabulary through the screening and arranging of texts. *Language Learning and Technology*, 6(1), 147-164.
- (27) Hitotsugi, I., & R. Day (2004). Extensive reading in Japanese. *Reading in a Foreign Language*, 16(1), 20-39.
- (28) Hu, M. H., & P. Nation (2000) Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- (29) Jeon, E. H., & J. Yamashita (2014) L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160-212.
- (30) Koda, K. (1989) The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Language Annals*, 22(6), 529-540.
- (31) Laufer, B., & G. C. Ravenhorst-Kalovski (2010) Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.

- (32) Long, M. (1991) Focus on form: A design feature in language teaching methodology. In De Bot, K., R. Ginsberg & C. Kramsch. *Foreign Language Research in Cross-cultural Perspective*. Amsterdam: John Benjamins, 39-52.
- (33) Martín, M. D., L. Jansen, & E. Beckmann (2016) Accounting for the missing students: Calculating retention rates in language & Culture programs. *The Doubters' Dilemma: Exploring Student Attrition and Retention in University Language and Culture Programs*. ANU Press, The Australian National University, Canberra, Australia.
- (34) Matsushita, T. (2012) In What Order Should Learners Learn Japanese Vocabulary? A Corpus-based Approach. PhD dissertation, Victoria University of Wellington. <<http://researcharchive.vuw.ac.nz/xmlui/handle/10063/4476>> (2017年8月19日参照)
- (35) Nation, P. (2007) The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2-13.
- (36) Nation, I. S. P., & J. Deweerdt (2001) A defense of simplification. *Prospect*, 16(3), 55-67.
- (37) Nation, I. S. P., & S. A. Webb (2011) *Researching and Analyzing Vocabulary*. Boston, MA: Heinle, Cengage Learning.
- (38) Norton Peirce, B. (1995) Social identity, investment, and language learning. *TESOL Quarterly*, 29(1), 9-31.
- (39) Richards, B. J., & D. D. Malvern (1997) *Quantifying Lexical Diversity in the Study of Language Development*. Reading: University of Reading.
- (40) Rott, S. (1999) The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, 21, 589-619.
- (41) Schmitt, N., X. Jiang, & W. Grabe (2011) The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- (42) Waring, R., & M. Takaki (2003) At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- (43) Yamashita, J. (2013) Effects of extensive reading on reading attitudes in a foreign language. *Reading in a Foreign Language*. 25(2), 248-263.

Text Modification in Japanese: Importance and Approach, with a Focus on Vocabulary Factors

MATSUSHITA Tatsuhiko

In recent years, the importance of text modification has been drawing attention. Firstly, this paper discusses the importance of text modification from the viewpoint of motivation, practicality, and the learning of reading and vocabulary. Relevant previous studies are overviewed concerning the following three points: 1) the importance of vocabulary in reading comprehension, 2) the relationship between text coverage and reading comprehension, and 3) the different appropriate levels of text coverage by known words for different purposes of reading, such as intensive reading, extensive reading, and speed reading. It is confirmed that text modification is important in order to adjust learning burden and enhance the chances of language acquisition. Secondly, this paper introduces web tools used for modifying texts and describes the rationale and method of text modification. Based on the aforementioned discussion, the paper proposes the Lexical Learning Possibility Index of Reading Text (LEPIX), and presents an example of text modification and the calculation of LEPIX. Finally, two remaining issues are pointed out, i.e., how we should quantify the lexical level of text and how we should incorporate the cognate effect into the approach for text modification.

【Keywords】 text modification, vocabulary learning, text coverage, lexical level, LEPIX