

首都大学東京都市教養学部人文・社会系  
首都大学東京人文科学研究科  
「人文学報」第512-7号  
(日本語教育学)  
2016年3月抜刷

語彙チェッカーを用いた  
語彙教材の分析とその問題点

長谷川 守 寿

# 語彙チェッカーを用いた 語彙教材の分析とその問題点

長谷川守寿

## 1. 目的

本稿は、語彙レベルを判定する語彙チェッカーを用いて、語彙教材を分析し、教材の改訂に役立てようという試みである。実際には、語彙チェッカーを用いることの利点、問題点を明らかにすることによって、今後の教材の語彙分析に生かしていきたいと考える。

教材作成について、国際交流基金（2008:iv）では以下の8つのステップがあるとしている。

- ステップ1：教材を作る前に
- ステップ2：教材を設計する
- ステップ3：企画書を作成する
- ステップ4：教材開発を進める
- ステップ5：試作し、評価する
- ステップ6：完成する
- ステップ7：使用し、改善点を明らかにする
- ステップ8：改訂する

関・平高（2015）では、教材開発のプロセスは、教科書開発のプロセスにも当てはまるとし、上記の国際交流基金（2008）と Jolly&Bolitho（2011:113）をまとめ、教科書開発の評価について、以下のように述べている。

出版後も教師、学習者の両者による評価が必要です。授業を通して使いやすさ、学習効果などのチェックを行い、必要であれば例文の差し替えなどを行って、増刷や改訂の機会があればそれに向けて準備をしておきます。

(関・平高 (2015,p.18)、一部表記を改めた)

しかし、ステップ7の「改善点を明らかにする」には、関・平高 (2015) が述べるように教師や学習者による評価も必要であろうが、出版後の教師による使いやすさや学習効果の評価は、使用している教師を特定するのが難しいため、ごく一部の教師に限られてしまうという問題がある。また学習者による評価は、授業で使用した感想を集めることで可能となるが、そもそもその教材で学習するため、全ての内容が理解出来るとは言いがたい学習者に評価が出来るのかという問題が残される。また、学習効果の評価は別として、教師や学習者による使いやすさの評価は漠然とした主観的な評価になってしまう恐れがある。

筆者は教材を作成する中で、執筆者の一人として他の執筆者（や時には出版社の担当者）とあてもないこうでもないと議論をする中で、例文を決定するプロセスを経て、教科書を作ってきた。改訂には網羅的にかつ一貫した作業が要求されるが、改訂の際に執筆者側として同様の人的資源を使用できるとは限らず、逆に割くことが出来ない場合が多いことが予想される。さらに、執筆者が教師として教科書を使用していく中で、改訂する必要がある問題点に全て気づけるわけではない。また、教師や学習者による、やや主観的になりがちな評価では、どこを修正したら良いのか結局執筆者の判断に任せられることになってしまう。そこで、筆者は客観的なツールによる評価を用いて、教材の改善点の候補を明らかにすることが出来るのではないかと考える。本調査では、客観的なツールとして、ある語が学習されるレベルが表示される語彙チェッカーを挙げる。これを用いれば、学習者にとって未習の語彙が判明し\*<sup>1</sup>、例文の差し替えなどを行うべき箇所の候補を、教科書の中で漏れなく挙げていくことが出来るのではないかと考え、語彙

---

\*<sup>1</sup> もちろん、教材に使用する語彙が、全て学習した語彙である必要はない。未習語彙でも、学習者の既習語彙から意味を推測することは必要である。しかし、改訂に向け、執筆者が再考すべき点を、漏れなく示すことは、改訂作業の支援になると思われる。

チェッカーを用いて教材内の語彙の評価を行う。そして、学習効果を使いやすさなどの主観的な観点からではなく、語彙の難易度という観点からの評価を行うことの利点と問題点を明らかにし、実際に教材を評価した結果を基に、差し替えなどを行うか検討すべき箇所を特定し、改訂の機会に向けた準備としたい。

## 2. 先行研究

日本語表現の難易度を明らかにするツールを語彙チェッカーと呼ぶこととする。語彙チェッカーの作成に関する研究としては、川村（1998）、松下（2014）、伊藤ほか（2014）、北村・孝橋（2015）等が挙げられ、「リーディングチュウ太」、「J-LEX」、「これやさしいか」、「かぶとエディタ」、「日本語文章難易度判別システム」（<http://jreadability.net>）等様々なツールが使用可能となっている。

実際にツールを使用して教材のレベル判定を行った研究には、川村（1999）があげられる。川村（1999）では、実際に「リーディングチュウ太」を用いて日本語教科書8冊（初級4冊、中級3冊、上級1冊）を対象に、語彙のレベル判定を行っている。教科書毎の1級・2級・3級・4級・級外の延べ語数は明らかではないが、初級・中級・上級と進むにつれて、1級・2級・3級・4級・級外の語彙数がどのように変化するか明らかにしている。ただし、川村（1999）は中級・上級に関しては判定の対象は読解教材部分のみであり、対象の文長を考慮して、初級は3課分、中級は2課分、上級は3課分と限定的であり、教科書全体の傾向を述べるには不十分であると思われる。

しかし、出版社から一般に入手可能な語彙教材自体が少ないことなどもあり、語彙教材に対するレベル判定に関する研究は行われてこなかった。現在では『耳から覚える日本語能力試験語彙トレーニング N3』（安藤栄里子・飯嶋美知子・恵谷容子著、アルク、2010年）『日本語総まとめ N3 語彙』（佐々木仁子・松本紀子著、アスク出版、2010年）など語彙教材自体は出版されるようになってきたが、教材が謳うレベルと実際に教科書で使われている語彙のレベルを確認し、教科書の真正性を明らかにするような研究や調査などは管見に限り行われていない。

### 3. 研究方法

#### 3.1. 対象

本調査では、小野正樹他（2009）『コロケーションで増やす表現—ほんきの日本語— vol.1』（くろしお出版）を対象に語彙のレベル判定を行う。この教材は、ある語（例えば「手」や「頭」）を中心とした語彙教材で、中心となる語は30語である。教材内では「課」という呼び方はしていないが、説明の便宜上30課とする。それぞれについて、語彙のレベル判定を行う。

筆者は本教材を授業で使用していく中で、実際の受講者に対して難しすぎるのではないかと感じる事が多くなった。昨今の学習者のレベルが、作成当時の学習者と比べて、相対的に低下しているということも考えられるが\*<sup>2</sup>、現状の学習者のレベルに相応しくない語を含んでいることは確かである。ちなみに、実際に使用している首都大学東京の未習言語科目（日本語I（上級））において2015年秋のクラスの単位取得を目標とした受講者はN1取得者が5名、N2取得者が1名である。授業のレベルを紹介したホームページには「日本語能力試験N1レベル」とされ、「アカデミックな日本語の基礎を総合的に学習します。」（[http://www.ic.tmu.ac.jp/study\\_abroad/jpclass.html](http://www.ic.tmu.ac.jp/study_abroad/jpclass.html)）とあるが、教科書が学生のレベルにあっていないという印象は否めない。出版から6年が経過し、授業で使用してくる中で、教材の中心となる語でもなく、難解で必要性のない語句の存在なども分かってきた。必要な部分に集中できるよう、分かりやすい語句に替え、支障が出ないようにする必要がある\*<sup>3</sup>。

#### 3.2. 手順

##### 3.2.1. 調査対象

出版社への入稿原稿であるpdfファイルからテキストを抜き出し、調査対象のデータとする。入稿原稿では答えを入れるために括弧になっている部分には、「別冊 解答」

---

\*<sup>2</sup> 実際に授業の最初の回に行う事前テスト（同一の問題を使用）の成績は、多少の変動もあるが、下降傾向にある。

\*<sup>3</sup> たとえば、「持つ」を中心に学ぶ課に含まれる例文「母にはリュウマチの持病がある」における「リュウマチ」などである。

から答えを入力した。なお、「手< >込んだ模様」のように複数の回答（「に／の」）がある場合は、初めの答えを入力した。また入稿原稿と製本は入稿後の校正により多少表記等が異なる部分があるが、その場合は製本を基準として修正する。

この本は以下の6部からなる。これらを指示文と問題番号を除きテキスト化した（さらに学習を深めるための課題「STEP 調べてみましょう」もあるが、実例ではなく指示が多く含まれているため除外した）。

- ・ 意味を記述した「こんな意味があります」（以後「意味」と呼ぶ）
- ・ 問題文の語がどの基本義に該当するか考える「基本語のどれに当たるか考えてみましょう。」（「どれ」）
- ・ 空欄に適切なことばを入れて意味を確認する練習（「問題1」）
- ・ 選択問題で該当する語と共起しやすい表現を学ぶ練習（「問題2」）
- ・ 該当する語の複合語などを学ぶ練習（「問題3」）
- ・ 今までに学んだ表現が新聞や小説の文体などで現れ、内容を確認する問題（「練習4」）

### 3.2.2. レベルの判定方法

本調査では、日本語の表現の難易度を判定するツールとして、「日本語読解学習支援システム リーディング・チュウ太」(<http://language.tiu.ac.jp>)（以後「チュウ太」と呼ぶ）を選択した。表現の難易度を判定するサイトやツールには、先行研究で挙げたように多々存在するが、広く長く使用されていること、インストールや辞書の準備などの必要がないこと、問題点も一部明らかになっていることなどその実績を考慮し、チュウ太を選んだ。

チュウ太のテキストボックスに各課のテキストを部分毎に入力し、「語彙レベル」のボタンをクリックする。結果画面は図1のように3つに分かれるが、本調査では画面右の級別の単語リストと、画面下の単語レベルの判定結果をコピーして使用する。その下にある「総数・語彙総数・級外」等をカウントした表は使用しない。なお、最新のチュウ太の頁には「Last updated on 2015-06-11」とあり、本調査は2015年1月15日に行ったため、現在でも同様の結果が得られる保証はない。

リーディング チュウ太

日本語 / English / Deutsch

---

**入力された文章**

頭でっちな人。  
ページの頭。  
頭上注意。

**級外 2 (2)**

頭でっちな 1  
頭上 1

1級 0

2級 0

3級 1 (1) 注意 1

4級 4 (4)

人 1  
ページ 1  
の 1  
頭 1

---

単語レベル ★★ やさしい									
総数	語彙総数	級外	1級	2級	3級	4級	その他		
10	7	2	0	0	1	4	3		
142%	100%	20%	0%	0%	57%	42%			
(8)	(7)	(2)	(0)	(0)	(1)	(4)	(1)		
114.3%	100.0%	20.0%	0.0%	0.0%	57.1%	42.3%			

図. 1 チュウ太の結果画面

チュウ太では「お知らせ」等が級外と判定されているという質問に答える形で、複合語・連語のレベル判定について以下のように説明している。

「語彙チェッカー」は入力された文章を形態素解析し、その結果を『日本語能力試験出題基準』（以下「出題基準」）の級別語彙リストと照合して単語のレベル判定をおこなっています。文の解析には、形態素解析システム「茶筌 2.02」（奈良先端科学技術大学院松本研究室が開発）を利用しています。そのため、「茶筌」と「出題基準」で単語の区切り方が異なっている場合、問題が生じます。例えば、上の例の「お知らせ」という語句は、「茶筌」はそのまま「お知らせ」と一語で区切ります。それに対して「出題基準」では（一般の国語文法もそうですが）、「お」と「知らせ」は別々の単語として扱っています。そのため「お知らせ」は「出題基準」のリストにないので、「級外」と判定されたわけです。（以下、省略）  
（川村（2009,pp.92-93））

上記のように、チュウ太には「茶筌」と「出題基準」で単語の区切り方が異なることから発生する問題が存在する。また、形態素解析システムも 100%正しく解析できるわけではない。そこで、本研究では、少しでもその問題点を解決するため、以下のような

手順で後処理を行うこととする。

### 3.2.2.1. 修正箇所の特定

筆者が一文ずつ解析結果である語の区切り方と語のレベル（級外・1級・2級・3級・4級・その他）と確認し、誤りがあれば修正する。修正が必要な場合は、類似する表現をチュウ太に入力し、その結果からどのように修正すればいいか確認する。類似する表現をチュウ太に入力しても同様の誤りが発生する場合は『日本語能力試験出題基準【改訂版】』（国際交流基金、財団法人日本国際教育協会・編、凡人社、2002年）を用い、語彙表から判定する（以後「出題基準」と呼ぶ）。

複合語の中には、チュウ太が使用している辞書によって、細かく区切られることがあるが、そのまま処理をする。たとえば、「バス代・バス通り・バスガイド・バス停」は、「バス（4級）代（3級）」「バス（4級）通り（2級）」「バス（4級）ガイド（1級）」「バス停（級外）」と判断される。「バス停」が級外になるのは、形態素解析システムの辞書に「バス停」は存在するが、級別語彙リストには存在しないためである。なおここでは「停」のみを級外とすることはしない。

### 3.2.2.2. 結果の修正

上記の結果に修正を加える。修正にはいくつかのタイプがあり、判定されたレベルから察して誤った読みで判定されているため正しい読みとそのレベルに修正する場合、区切り方に誤りがあり統合してそれにふさわしいレベルを再付与する場合などがある。これについては結果で詳細に述べるが、たとえばチェーン店は「チェーン／店」に分けられ、「チェーン（級外）／店（4級）」の語と判定される（以後“/”は語の区切りを示す）。しかし「店」で4級と判定されるのは読みが「みせ」の場合だけで、「てん」は2級と判定されなければならない。これは「特定が難しく、合致する語彙が複数候補考えられる場合には、下の級を優先」（川村1999）したことによるものと思われる。このような場合は修正が必要である。

また機能語などは、語彙表に掲載がないため、級外と判断されてしまう。たとえば、「によって」「に対する」は、形態素解析辞書では一語とされ、語彙リストとのマッチング



が出来ないため級外と判定されるが、このような表現の場合、「出題基準」の「Ⅰ. 文法(3・4級)のリストについて」と「Ⅱ. 1・2級の<機能語>の類」のリストから判定し、ともに2級のサンプルに含まれているため、これらは2級に修正する。

なお、形態素解析では1語ではなく複数に分割され、それぞれに級が判定される場合は、そのまま判定された級を用いることとする。たとえば「ねばり腰」は1語ではなく、「ねばり(1級)／腰(2級)」と2語として判断されるが、統合することなくそのまま集計する。

### 3. 2. 2. 3. 結果の集計

30課を通した単語レベルの判定、6つの部分毎の30課を通した基本統計量、30課を通した語彙レベル毎の延べ語数をまとめる。

具体的には、各課の「意味」「どれ」「1番」「2番」「3番」「4番」の単語レベルの判定を1課から30課まで集計し、判定毎にまとめる。

次に1課から30課まで、各課に含まれるテキストの延べ語数を、「意味」「どれ」「1番」「2番」「3番」「4番」に分けて、集計する。

最後に、各課に含まれる全てのテキストをチュウ太にかけた結果を、級外・1級・2級・3級・4級に分けて延べ語数を集計する。

なお川村(1999, 2009)などには記述がないが、チュウ太には一度に入力できるテキストの量には制限があるようで、全てのデータを入力すると、「入力された文章が長すぎます。短くして再度入力してください。コード:6」という表示が出てしまい、解析結果が途中までしか表示されない。そのため、「意味」「どれ」「2番」「3番」については、1課から30課までのテキストを全てまとめて解析することが出来、延べ語数・異なり語数を得ることが出来たが、文字数の多い「1番」「4番」は、全てのテキストを入力すると結果が得られず、異なり語のリストとそれぞれの頻度については、取得することが出来なかった。そのため全体の集計に関しては延べ語数のみを対象とする。

## 4. 結果

まずはチュウ太の結果に加えた修正点を示す。これはそれ以降の集計に関わるため

ある。その後、修正を加えた数値を示すこととする。

#### 4. 1. 修正点

チュウ太の結果に加えた修正点を示す。これはチュウ太の問題点を示すことでもあり、今後チュウ太を使用して語彙レベル判定を行う際には必要となる情報である。多く見られた誤りとその原因、その修正の方法について説明する。全ての誤りと修正を載せる紙幅はないため特徴ごとにまとめて示す。

##### 4. 1. 1. 読みが正しく同定されていない問題

まず、多く見られた誤りとして、漢字の読みが正しく同定されないことによる誤解析がある。「特定が難しく、合致する語彙が複数候補考えられる場合には、下の級を優先」(川村 1999) したことが原因であるが、単純に下の級に判定されていないものもあるため、文脈から判断して修正を行う。

たとえば「象が十頭いる。」の波線部に対して、チュウ太は「頭」を1級としている。1級で「頭」と記載されているのは「かしら」の読みの場合であるが、この文での読みは「とう」なので正しくない。この場合、頭(とう)は2級の語なので2級に修正する。同様に「勝ち頭」の頭(がしら)を4級(あたま)と解析しているが、この場合は1級(かしら)に修正する。以下に同様の例を示すが、括弧内左の読みと級を、括弧内右の読みと級に変更した(なお、今回は漢字の読みを考慮した異なり語数の集計は行っていない)。読みが正しく同定されていない例としては、案内所(ところ・4級=>じょ・2級)、政治家(いえ・4級=>か・2級)、京都市(いち・1級=>し・2級)、留学先(せん・1級=>さき・2級)などがある。チュウ太を使用する際には、読みが多数ある漢字は以上のような誤った判断が見られるので、注意が必要である。

##### 4. 1. 2. 文が正しく語に区切れていない問題

語の区切り方が正しくないのは形態素解析システムの問題であるが、たとえば、「あんな石頭」の「あんな」は、「ある(4級) / だ(4級)」となる。調べてみると「あんな人」でも同じ結果が生じるため、「石頭」の問題ではない。「出題基準」の語彙リスト

を参照し、この場合は「あんな」一語に統合し、さらに3級に修正する。

他に語の区切りに誤りが起こる例として、動詞を誤って抽出する場合がある。たとえば、「本日は頭金だけお支払い下さい。」の場合（「だけ」の後に「お」が含まれていることにより発生するようである）、チュウ太は「だけ」の部分を「だける」（2級）の活用形と判断している。このような誤解析に対して、「頭金だけ支払いました。」では「だけ」は4級と判断されるので、同様に4級に修正する。

また動詞が正しく抽出できない場合もある。たとえば「いい顔ばかりもしてられない」に対しては、「いい／顔／ばかり／もし／て」のように語に分割し、「もし」は基本形を「もす」と解析し、級外に分類しているが、これも誤った語の切り方なので「も（4級）／する（4級）」と修正する。

#### 4. 1. 3. 接続規則・辞書・語彙リストに起因する問題

茶釜の接続規則に原因があると考えられる例がある。たとえば、「コンセプトはいい線行ってるけど」内の「いい」は、「いう（4級）」と解析されてしまう。ただし、「いい線行ってるけど」だけならば「いい（4級）」と判断されるため、「は」を含む接続規則による影響と思われるが、このような場合は正しい語形である「いい（4級）」に修正する。

また、辞書に登録されていないことが原因で、動詞の誤解析が起こるパターンがある。たとえば「起きたてで」は、「たて」が1級と判断されるが、1級で「たて」という読みを持つものは「盾」しかなく、機能語リストにもない。「作りたて」「炊きたて」は辞書に登録されているため、一語として「級外」と判断されているが、「できたて」「焼きたて」は同様に「たて」のみが1級と判定される。そこで、このような場合「起きたて」のように一語に統合し、「級外」に修正した。

また、辞書の情報が不十分で語彙のレベル同定が誤っている例がある。例えば、「筆記試験より、面接試験の方が緊張する」の波線部「より」を2級と判断している。しかし「出題基準」の「文字・語彙（1・2級）」の「Ⅲ. 語彙」の部分には「より」「よりいっそう」と記載されていて、正しい級が判断されているとはいえない。この場合「I. 文法（3・4級）のリストについて」（p.134）の例文「あの店よりこの店のほうがいい

です」などを参考に格助詞の「より」は3級に、副詞の「より」は2級になるように修正する。

語彙リストに起因する問題としては、同音異義語の例が挙げられる。たとえば、形容詞では「暖かい」は4級と判定されるが、「温かい」が級外と判定されてしまう。これは川村(2009)で説明があった「語彙リスト」に含まれていないためであり、出題基準を参考に適切な級に修正する。

またチュウ太では、数字やアルファベットだけのものは「その他」と判定される。たとえば、「300」や「CM・DVD・OB・PC」は句点や読点と同様に「その他」に分類されるが、これも辞書なり語彙リストの問題と考えられる。本調査では上記のようなアルファベットからなる語は級外に修正した。数字は、「出題基準」の「日本語能力試験「出題基準」(文法) - 4級-」の「6 a) 数の言い方」(p.128)を参考に、4級に修正した。

また辞書に関連する問題で、ひらがな表記された語と漢字表記された語の判定が異なる場合がある。たとえば、「ものごと」は「もの(4級) / ごと(2級)」と判断される。この場合は、「毎」の意味を持つ「ごと」を2級の語としているためであるが、漢字表記した「物事」は2級と判定されるので、一語に統合し、2級に修正した。

なお、本調査では異なり語数に関する調査を行わなかったため影響はないが、表記に関連する問題も見られた。たとえば「これは良い本だ。これはいい本だ。これはよい本だ。」は「いい・よい・良い」と別の語として出力されるため、複数の表記が可能な場合は注意が必要である。

#### 4.1.4. その他の問題

まず「出題基準」の問題に触れておく。2010年から実施されている新しい日本語能力試験からは出題基準が非公開になったが、「出題基準」は今でも多く使われているため、問題点を示しておくことは一定の意味があると思われる。処理に困ったのが、基準が複数ある場合であり、たとえば2級の語彙リストに「\*～か(家)」(p.60)があり、3級の語彙リストに「～か(家)」(p.23)がある。これは前出の「店(みせ)」(4級)・「店(てん)」(2級)のように、読みによる判定が出来ないという問題が発生する。このような

場合は、川村（1999）に倣い低い方の判定に合わせた。

また、語彙のリストにはなく文法項目から判定せざるを得ない際にも問題が生じる。たとえば終助詞の「な」については「(文法) - 3級-」に掲載されているが、禁止の用例しか示されていない。終助詞には詠嘆など他の用法も含まれるが、他の用例が掲載されていない。このような場合は、チュウ太の結果を優先した。

なお、最後に本調査に現れたチュウ太の問題点を記しておく。これは上記の問題とは違い、全てのテキストに当てはまるとはいえないのであるが、語と語が改行で区切られている場合、つまり句点「。」で区切られていない場合)、一語としてしまうことである。たとえば、ある文が「大」で終わり（句点「。」はなく）、次の行が「家に話を聞く」のような場合、「大家（1級）」と判断してしまう。教材には箇条書きなども含まれると思うが、そのような項目をチュウ太にかける際には、あらかじめ句点を入れておく等、対策が必要となる。

## 4.2. 集計結果

### 4.2.1. 単語レベルの判定

チュウ太では、単語レベルの評価として、「とても難しい」「難しい」「少し難しい」「やさしい」という判定を下している\*<sup>5</sup>。表1は各項目に対してチュウ太が判断したレベルをまとめたものである。意味の部分は「とてもやさしい」と判断された課が2課あることを示す。

---

\*<sup>5</sup> 川村（2013）では、単語レベルは「文に含まれる単語の難易度をもとに、入力文の難易度を判定したもの」（p.202）で、「難易度の判定は、総単語数に占める3級と4級の単語の割合を基準に」（p.202）、「とてもやさしい」（80%以上）、「やさしい」（70%以上80%未満）、「ふつう」（60%以上70%未満）、「少し難しい」（50%以上60%未満）、「難しい」（50%未満）と表示している

表1. 単語レベルから見た各問題の結果

	意味	どれ	1番	2番	3番	4番
とてもやさしい	2	1	0	4	0	0
やさしい	1	7	5	11	3	7
ふつう	10	9	19	14	16	10
少し難しい	15	9	6	1	11	12
難しい	2	4	0	0	0	1
総計	30	30	30	30	30	30

「どんな」では「少し難しい」「難しい」の合計が30課中13課で、やや難しい部分かと思われる。「1番」は「ふつう」が30課中19課で、「やさしい」「少し難しい」が5課と6課で、「ふつう」が多く、理想的な分布ともいえる。語数が多いためにこのようになっているとも考えられる。「2番」は、「とてもやさしい」「やさしい」の合計が15課、「ふつう」と「少し難しい」の合計が15課となっており、多少「やさしい」側に偏った分布かと思われる。「3番」は、逆に「少し難しい」が11課となっており、多少「難しい」側に偏っていると思われる。「4番」は最多が「少し難しい」であり、「難しい」側に偏っている。これは新聞や小説風の文体で使用例を統一しているため、このような結果になったと考えられる。「意味」は「少し難しい」が最多となっており、これは箇条書きの要領で語の意味を短く記述しようとしたため助詞が少なく「難しい」と判断されたものと推測される。

川村(2009)には、単語レベルと学習者のレベルに関して以下のような説明がある。

「ふつう」の表示は、上級学習者(2級合格以上)から見た文章の難易度を表します。(「とてもやさしい」「やさしい」「ふつう」「少し難しい」「難しい」の5段階です。)中級の学習者にとっては、「ふつう」の文章では難しく感じられるため、「やさしい」と表示される程度の教材を選ぶ必要があります。(川村(2009,p.19))

現在の日本語能力試験の結果は、2009年までの試験(旧試験)と以下のように対応

している (<https://www.jlpt.jp/about/comparison.html> より)。箇条書きに改め、句点を追加した)。

- ・ N1 合格ラインは旧試験の1級とほぼ同じ。ただし、旧試験の1級よりもやや高めのレベルまで測れるようになる。
- ・ N2 旧試験の2級とほぼ同じレベル。

N2が2級と同様と考えると、N2の学生にとって、「少し難しい」「難しい」が3分の1以上を占める「意味」「3番」「4番」は難しく感じているものと思われる。ここに原因があるのではないかと思われる。

個別に観察すると、級外と1級の延べ語数が多いと「難しい」と判断され、級外と1級の延べ語数が少ないと「とてもやさしい」と判断されるようで、級外の延べ語数が判断に大きく関与しているようである。そこで級外の語彙については別に章を設け、詳述することとする。

#### 4.2. 全体の集計

1課から30課までの、部分別の基本統計量を表2に示す。

表2. 30課全体の部分別基本統計量

	延べ語数	比率 (%)	一課あたりの平均語数	最大値	最小値	中央値	標準偏差
意味	689	3.5	23.0	32	9	23.5	6.1
どれ	1247	6.2	41.6	58	26	41	8.8
1番	6999	35.1	233.3	286	181	234	25.9
2番	2763	13.8	92.1	118	68	90.5	12.3
3番	2554	12.8	85.1	128	66	80.5	15.4
4番	5710	28.6	190.3	262	139	188	27.1
合計	19960	100.0	324				

1番の問題が全体の語彙数の35%程度、4番の問題が30%弱、語を選択肢から選ぶ

問題である2番と3番は多少2番が多いがそれほど変わらない結果となった。標準偏差などを見ても、紙幅の影響もあって、それほど語数にばらつきがないという印象がある。

今後の予定でも述べるが、これは教材1つの情報であり、今までこのような研究は行われてきていない。この結果と対照するデータがないため、これと比較するデータを今後求めていく必要がある。

### 4.3. 級別の集計

全ての教科書のテキストに対して、記号など「その他」に分類されるものを除いた「級外」「1級」「2級」「3級」「4級」「語彙総数」をまとめた結果を表3に示す。

表3. 級別の延べ語数の基本統計量

	延べ語数	比率 (%)	一課あたりの平均語数	最大値	最小値	中央値	標準偏差
級外	1600	8.0	53.3	76	38	53	9.6
1級	600	3.0	20.0	32	8	19.5	6.4
2級	2404	12.0	80.1	130	64	76	14.8
3級	2246	11.3	74.9	106	54	75.5	13.8
4級	13110	65.7	437.0	540	325	434.5	54.7
計	19960	100.0	665.3	787	555	663	53.5

表3から分かるように、4級の語彙の延べ語数が66%近くを占めている。これは初級の助詞などが4級に分類されるためであると思われる。また3級と2級の語彙の延べ語数はともに11～12%である。「上級レベルの学習者」を謳っている（小野他2009, iii）割には、1級の語彙の延べ語数は全体の3%と少ない印象を受ける。級外が8%でこれを全て未習語彙と考えると、2級までの学習者にとっては、1級と級外の語彙が1割を超え、1級取得者にとっても未習語彙が8%ほど出現していることになる。そこで級外の語彙について次の章で詳述する。



#### 4. 4. 級外の語彙について

級外の語彙は未学習であると仮定すれば、これらの語彙を改訂の際には検討する必要があるのではないと思われる。そこで、まずどのような語が含まれるのか、観察を行う。表4は、級外の語彙に対し、延べ語数の多いものを頻度5までの語を取り出したものである。

表4. 級外語彙とその頻度（頻度5以上）

頻度	語
26	けど
16	締まる
14	利く
9	なあ
8	プロジェクト
7	容疑
6	社員、思い、ぞ、って
5	肌寒い、知らず、口調、顔色、なあ

表4に含まれる「利く」は、同音異語である「聞く」が4級、「効く」が2級に含まれるが、「利く」の出現形は語彙リストに入っていない。また「締まる」は、2級に他動詞である「締める」は含まれているが、自動詞形は含まれていない。これらの語は、類義表現が含まれていない「出題基準」の問題ともいえる。しかし、表4には口語的表現が多く含まれ、表記の揺れを合算すれば、「けど (26)」「なあ (14)」「って (6)」が挙げられる。口語的表現がこれだけ例文中に含まれていたのであるが、口語的な例文を含んでいいのか、再考が必要となろう。他に表4には「プロジェクト」が8回出現しているが、カタカナ語では「コンビニ・セールスマン・ボクシング・ローン」それぞれ3回出現し、カタカナ語の多さも指摘できる。

また、級外語彙は延べ語数が1,600語で、異なり語数は表記の違いを考慮し統合を行った結果1,201語であった。これらの語種を調べた結果が表5である。語種の決定には

unicid-mecab 2.1.2 と茶まめ Ver.2.0 を用いた。

語種を調べるにあたり、辞書が異なるため語の区切り方が異なるという問題が発生する。チュウ太が使用している形態素解析システムの辞書では、たとえば「時間切れ」は一語であるが、茶まめでは「時間（漢語）／切れ（和語）」となる。そこで結果について全て筆者が確認し、必要があれば修正・統合し、この場合「時間切れ」は、一語で語種は混種語とした。同様に「カメラブレ」「Tシャツ」は、「外来語／和語」「記号／外来語」と解析されるため、一語に統合し混種語とした。なお「固有」には、地名・人名が含まれるが、「朱（あか）」「高山（こうざん）」「山中（さんちゅう）」も「固有」に含まれていたため文脈に合わせ修正した。

級外の語彙においては、和語が半数近くを占めている。和語を精査すると、「乗り越える・乗り越す・乗り回す・乗り捨てる・乗り出す・乗り上げる・乗り切る」のような複合動詞や、「身だしなみ・身軽・身元・身動き・身売り」のような複合名詞が多く見られた。これらの語は、選択肢の中に挙げられ、変更の困難な語句であろう。

しかし問題4では新聞・小説の文体をまねて例文を作ったため、固有名が多い。たとえば、「青森県」「長野県」「お台場」「日光」等特定の地域名や「徳川」「鷗外」等の歴史に関連する個人名なのである。これらが目標とする語の学習に必要なのか、再検討する必要がある。

さらに検討の余地がある語についてみていく。和語には「ガラガラ・キラキラ・ゴットン・ころころ・ちりりん・バリバリ・ポロポロ・ムンムン」等のような擬音語・擬態語が含まれる。外来語には「プランクトン・ホスピス・リュウマチ・レシピ・ペンション」など、記号には「OB」「CM」などが含まれる。これらは学習者にとっては理解できない語である場合が多く、語だけでなく使用する文全体の変更も含め、検討が必要である。

級外の語彙に対しては、その目的は異なるが「減災のための「やさしい日本語」」の中の「「やさしい日本語」にするための12の規則」(<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ9tsukurikata.ujie.htm>) が、一部有効なのではないかと思われる。規則には「(4) カタカナ・外来語はなるべく使わないでください」「(6) 擬態語や擬音語は使わないでください」があり、「外来語は原語と意味や発音の異なるものが多い

ため」(同上)、「日本語で使用される擬態語(めっちゃめっちゃ、どきどきなど)や擬音語(ドンドン、ガシャンなど)は、日本語話者以外には伝わりにくい表現」であるため、避けるべきとしているが、この理由はそのまま学習者にも適用でき、教科書も改訂にも役立つものと思われる。

これらの語が理解できないために、その課で目標とする語を含む文の意味が理解できないことがないよう、必要な語なのか改訂の際には検討する必要がある。

## 5. 考察

ここでは、チュウ太を用いることのメリットについて述べる。

まず、教材の著者の一人としては、表記の揺れを確認できたことに驚かされた。例えば「充分」と「十分」、「気づく」と「気付く」、「10分」と「十分」等、筆者や編集者の目を通して漏れてしまった表記の揺れが見つかった。表記の揺れが離れた場所に存在する場合、校正者であっても気づきにくい問題であろう。このようなツールを用いることにより、表記の揺れが明確になった。これは特に複数の筆者によって執筆されている場合などに有効なものではないかと思われる。また、チュウ太では句読点の有無によって、形態素解析に結果が変わるという特徴がある。たとえば「コンサートで、」という部分では、波線部を断定の助動詞「だ」の連用形として解析しているが、読点「、」がなければ格助詞として解析する。このように不要な読点の存在に気付かされ、ツールを通した結果を考察することは、表記の統一にも一役買うのではないかと思う。また、「パンツのチャックが開いていたことに気がついて」という例文中の「チャック」は「ファスナー」の商標名であるが、そのまま教材に掲載されていた。このように級外と判断されることによって、使用されている語の問題に初めて気がつくという面も存在する。

また、4.4の「級外の語彙について」で言及したことであるが、集計し数値化することにより、教材の中では特に目立った印象がない話し言葉的な表現の存在が明らかになった。ここは改訂の際の修正点になり得ると思われるが、そもその問題として、習得させたかったのは書き言葉でのコロケーションなのか、話し言葉でのものなのか明確ではなかったと思われる。この教材には話し言葉と書き言葉が混在しており、教材の設計の段階でどのジャンルにおける用法か、統一した方針を明確にしておく必要があった

と思われる。

本稿では教材の評価にチュウ太という語彙チェッカーを用いた場合の利点と問題点を述べてきた。チュウ太の本来の使用方法が、学習者が読解教材を自習する際の支援であったり、教師が教材を準備する際の支援であるとすれば、本来の使用方法とは異なるが、少なくとも上記のような項目は改訂の機会があれば修正すべき点であり、このような点に目を向けさせてくれるという意味で有効な手段たり得るのではないだろうか。

## 6. 今後の予定

本調査は試行的なもので、同様の調査を行ったものがなく比較の対象がない。そのため、今後は他の語彙教材を対象とした調査が必要である。今回、同時に複数の語彙教材を対象として調査を行うことも考えたが、語彙チェッカーそれ自体の特徴を明らかにするために、教材を一つに絞らざるを得なかった。今後はこの方法を用いて、別の語彙教材との比較研究を行い、改訂に関する実証的な考察を行っていきたい。

## 参考文献

- 安藤栄里子・飯嶋美知子・恵谷容子 (2010) 『耳から覚える日本語能力試験語彙トレーニングN3』、アルク
- 伊藤恵・伊藤美紀・藤田篤・木塚あゆみ・大塚裕子 (2014) 「日本語支援者支援システムの構築－日本語教員養成を題材として－」『情報処理学会研究報告』 vol.2014-CE-125、No.5、情報処理学会
- 小野正樹・小林典子・長谷川守寿 (2009) 『コロケーションで増やす表現 ほんきの日本語 vol.1』、くろしお出版
- 川村よし子 (1998) 「語彙チェッカーを用いた読解テキストの分析」、早稲田大学日本語研究教育センター、『講座日本語教育』、第34分冊、pp.1-22
- 川村よし子 (1999) 「語彙チェッカーを用いた日本語教科書の分析」、The Second International Conference on Computer Assisted System for Teaching & Learning Japanese CASTEL/J'99 Proceedings、pp.132-137
- 川村よし子 (2009) 『チュウ太の虎の巻－日本語教育のためのインターネット活用術』、

くろしお出版

川村よし子 (2013) 「リーディング・チュウ太と「やさしい日本語」、庵功雄・イヨンスク・森篤嗣編『「やさしい日本語」は何を目指すか 多文化共生社会を実現するために』、ココ出版、pp.199-217

北村達也・孝橋一希 (2015) 「文難易度の調整を支援するシステム「かぶとエディタ」」『日本語教育方法研究会誌』22 (1)、pp.14-15

国際交流基金 (2008) 『国際交流基金日本語教授法シリーズ第14巻 教材開発』、ひつじ書房

国際交流基金・財団法人日本国際教育協会 (2002) 『日本語能力試験出題基準【改訂版】』、凡人社

佐々木仁子・松本紀子 (2010) 『日本語総まとめ N3 語彙』、アスク出版

関正昭・平高史也 (2015) 『日本語教育叢書つくる 教科書を作る』、スリーエーネットワーク

松下達彦 (2014) 「オンライン日本語テキスト語彙分析器 J-LEX」『日本語教育方法研究会誌』21 (1)、pp.8-9

Jolly, D & Bolitho, R (2011) A framework for materials writing. In: Tomlinson, B. (ed) *Materials Development in Language Teaching*, Second Edition, pp.107-134. Cambridge: University Press.

## 参考サイト

旧試験との比較 日本語能力試験 JLPT

(<http://www.jlpt.jp/about/comparison.html>. 2016年1月31日最終確認)

日本語授業 - 首都大学東京国際センター

([http://www.ic.tmu.ac.jp/study\\_abroad/jpclass.html](http://www.ic.tmu.ac.jp/study_abroad/jpclass.html). 2016年1月31日最終確認)

日本語文章難易度判別システム

(<http://jreadability.net/>, 2016年1月31日最終確認)

「やさしい日本語」にするための12の規則

(<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ9tsukurikata.ujie.htm>. 2016年1月31日最終確認)

