日本語学習者用の日本語・スロベニア語辞書のための 読みやすいアクセント表記の試み

寒川仁衣菜（リュブリャナ大学）[†]
ヤン・フラストニク（リュブリャナ大学）
クリスティナ・フメリャク寒川（リュブリャナ大学）

An attempt at legible accent notations for a Japanese-Slovene learners' dictionary

Nina Sangawa Hmeljak (University of Ljubljana)
Jan Hrastnik (University of Ljubljana)
Kristina Hmeljak Sangawa (University of Ljubljana)

要旨

強弱アクセントの言語を母語とする学習者にとって、日本語の高低アクセントの認識と発声は難しい。高低差が聞き取れない学習者にとって、目に見えるアクセント情報の表示が正しい発音の助けになると考え、日本語・スロベニア語辞典の改訂にあたって、アクセント情報も付与することにした。そこで、「日本語教育語彙」から入手したアクセント情報を見出し語に付与したが、数字などの記号によるアクセント情報は学習者にとって分かりにくいと判断し、辞典の見出し語を初級日本語学習者にも読みやすいローマ字に変換し、高低アクセントを大文字と小文字、上付き文字、色分けなどで表した。本発表では、その過程を説明し、パイロット版の資料を利用した学習者の反応を紹介する。

１．はじめに

強弱アクセントの言語を母語とする日本語学習者にとって、高低アクセントは困難な学習項目の一つである。その概念を理解しても、母語にない音の高低のわずかな差を聞き取るのも、記憶するのも、発声するのも難しい。

新しい単語を習うときに、発音と同時にアクセントも覚えるのが望ましいのだが、多くの教科書には、漢字とカナやローマ字表記があっても、アクセントの情報が表記されていないものも多い。例えば、多くの教育現場で使われている教科書シリーズ、3A ネットワークの『みんなの日本語』、ジャパンタイムズの『げんき』、くろしおの『とびら』など、語彙の音声ファイルがあるものの、語彙リストの活字にはアクセントの情報はない。

また、日本語学習者が単語を調べるのによく使うアプリ（フメリャク寒川 2019、石黒・吉・佐野 2024）の多くは JMdict (EDRDG 2024) のデータを利用するが、そのデータにもアクセントの表記はない。単語の音声を再生できる辞書やアプリもあるが、聴覚では高低アクセントが聞き取れない、あるいは聞き取った音の判断に自信がない学習者にとって、聴

[†] nina.sangawa@gmail.com

覚と平行して視覚の情報も学習の支援になりうると考えられる。

　そこで、以前リュブリャナ大学で開発した日本語· スロベニア語辞書 jaSlo (Hmeljak et al. 2016) の改訂にあたって、見出し語にアクセントの表記も加えることにした。以下では、表記方式の種類と特徴について述べ、用意したデータを紹介し、未解決の課題に言及する。

２．アクセントの表記方式

　日本語標準語アクセントの表記方式は、概ね２グループに分けることができる。一つ目は、低く発音される拍と、高く発音される拍を、何らかの記号か視覚的な特徴で書き分けて表記する方式である。例えば、神田 (2003:53) にまとめられているように、

　① - ② 高く発音される拍の上に線を引いて、下り目を ①鍵で示すか、②示さないか、

　③ 上がり目も下り目も鉤括弧で示し、高く発音される拍の上に線を引く、

　④ 高音の上と低音の下に線を引いて、上がり目と下り目を縦の線で表す、

　⑤ 高く発音される拍を太文字で表記する、

　⑥ - ⑨ 単語の文字と平行して、上下を表す記号（上下、LH、●○など）、または上下に配列され、棒でつなげられている丸（○）で、アクセントパターンを示す、

　⑩ ローマ字表記に鉤括弧「 ⌉ で上がり目と下り目の境目を示す、

　⑪ 楽譜のように、平行する２線と重ねて、高音を上の線に、低音を下の線に書く、

　⑫ 高く発音される拍の上に傍点をつける。

　⑬ 図１にはないが、高い拍を別の色で書く、などの方式がある。

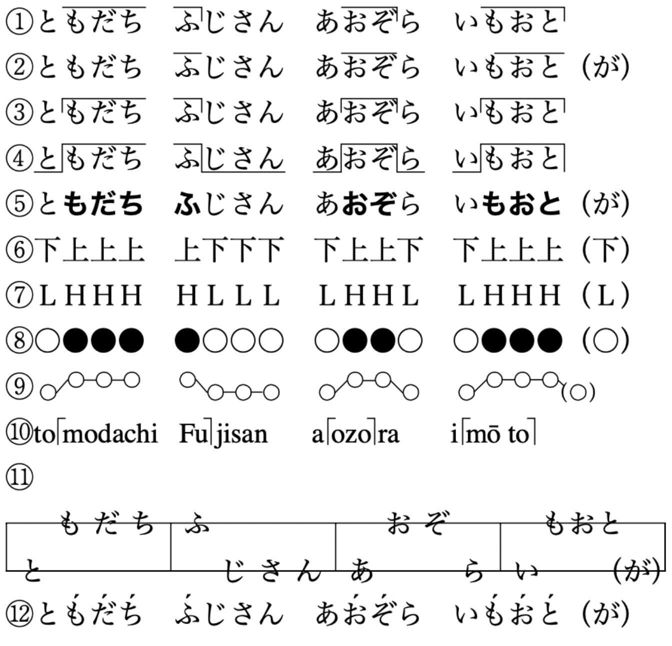

図 1 アクセント表記方式の例（神田 2003:53 による）

二つ目の表記方式のグループは、下り目の直前の拍、つまりアクセント核をなんらかの形で記す方式である。例えば、表１のように、

①単語の何拍目がアクセント核（下り目の直前の拍）かを数字で表す、
②アクセント核の文字の右上に鉤括弧をつける、
③アクセント核の右に斜め線（＼）をつける、
④アクセント核を目立つ色で書く、
⑤ローマ字表記の場合、アクセント核の母音にアクセントをつける、などの方式がある。

表１の④は OJAD で採用されている方式だが、それぞれの拍の高低と、アクセント核を、両方示しており、表記方式のグループ１とグループ２の組み合わせである。

表 1　アクセント核のみを記すアクセント表記の例

①	ともだち [0]	あいさつ [1]	おととい [3]	いもうと [4]
②	ともだち	あ˥いさつ	おとと˥い	いもうと˥
③	ともだち	あ＼いさつ	おとと＼い	いもうと＼
④	ともだち	あいさつ	おととい	いもうと
⑤	tomodachi	áisatsu	ototói	imoutó

京阪式アクセントでは、下げ核の他に、語頭の高低も弁別されるが、共通語（東京式）アクセントでは、１拍目と２拍目の間に必ず高低の差があり、２拍目以降が低い頭高の語以外、１拍目が低く２拍目は高いことになるので、下げ核につける印一つで単語の全拍の高低がわかる。つまり、上記の一つ目の表記方式グループより表記が経済的だが、初級学習者にはわかりにくい可能性もある。

３．データの装備

インプットデータとして『日本語教育語彙』（JEV 2021) から入手した 17948 の単語のカタカナ表記を使ったが、単語の「読み」として掲載されているカタカナは、標準かな表記であり、長音は「ケイタイ」、「コウギ」などのように記されているので、そのリストをまず、現代 UniDic を使った Web 茶豆で処理し、単語の発音形も得た。

次に、カタカナで書かれたそれぞれの単語の発音形を一つの txt ファイルに、そしてそれぞれの単語のアクセントを表す数字をもう一つの txt ファイルに保存した。アクセントを表す数字は０から９まであり、それぞれ以下の意味を表す。

０の場合、一番最初の拍が低く発音され、続く拍は全て高く発音される。
１の場合、一番最初の拍が高く発音され、続く拍は全て低く発音される。
２の場合、一つ目の拍が低く、二つ目が高く、その次の拍は全て低く発音される。
３の場合、一つ目の拍が低く、二つ目と三つ目が高く、続く拍は全て低く発音されて、

４以上の数字も同じように、一つ目の拍が低く、二つ目から表記の数字の拍まで高く、その後の拍は全て低く発音されることを表している。

表２は、『日本語教育語彙』におけるそれぞれのアクセント型の単語数を表す。

表 2『日本語教育語彙』のアクセント型別の単語数

アクセント型	0	1	2	3	4	5	6	7	8	9
単語数	8876	4214	1540	1869	903	447	72	21	5	2

上記の数字での表記は辞書でよく使われるが、初心者への分かりにくさを改善する為に、Python のプログラムを使って、より読みやすい五つのアクセント表記へ変えてみた。

最初にカタカナ表記を二つ用意した。一つ目は、リュブリャナ大学の初心者用日本語教科書 (フメリャク他 2012) で使われている表記である。発音の高低に差がある二つの拍を表すカタカナの間に、高く発音される拍から低く発音される拍の間には【⌉】の表記を挟み、低音から高音に変わる境目には【⌈】の表記を挟む方法である。図２にその例を示す。

	な⌈くしま⌉す	な⌈くさな⌉い	な⌈くして	V5	izgubiti
返します	か⌈えしま⌉す	か⌈えさ⌉ない	か⌉えして	V5	vrniti

図 2 リュブリャナ大学出版の『初級日本語１』、語彙表の抜粋

この表記を一番最初に処理した理由は【⌈】と【⌉】の記号を区切り符号としてのちのデータ処理にも使える為である。

最初に上記のカタカナで書かれた発音形と数字で表したアクセントの txt ファイルをそれぞれ Python のリストにし、単語のリストを一単語ずつ繰り返し処理した。まず単語を拍に分ける為に、最初のふた文字がカタカナの大文字と小文字の場合、小文字が「ッ」でなければ一つのエレメントとし、それ以外の場合は最初の文字を一つのエレメントとした。他の文字も順番に同じようにエレメント（拍）に分けた。文字がカタカナの小文字かどうかを確認するために Python の「unicodedata」を使ってそれぞれの文字の Unicode 名が「KATAKANA LETTER SMALL」で始まるかどうかを確認し、分別した。次にプロセスしている単語のインデックスと同じインデックスにあるアクセントの数字表記リストにある数字によってアクセントを表す記号を足していった。

アクセント型が０の場合、拍のリストのインデックス１（一番最初の拍と二番目の拍の間）に【⌈】を追加した。アクセントが１の場合、リストのインデックス１に【⌉】を追加した。アクセントが２以上の場合、アクセントの数字を「n」とする。拍のリストのインデックス１に【⌈】を追加し、インデックス「n」に【⌉】を追加する。最後にアクセントの記号が入ったリストを一つの文字列へマージし出力した。

二つ目の表記方式は『日本語発音アクセント新辞典』(NHK 2016) で使われている方法で、

高く発音される拍から低く発音される拍への境目に【＼】という記号を挟み、０のアクセント、つまり二つ目の拍から最後まで高く発音される場合は言葉の最後に【￣】という記号を入れる表記方式である。これは、共通語では一つ目と二つ目の拍の間に必ず高低の差があり、二つ目以降の拍の発音が一度下がれば二度とあがらない法則を利用した方法である。

最初の変換に使ったアルゴリズムとほとんど同じやり方で変換した。【˥】の代わりに【＼】を追加し、０型で単語が一拍の場合、【˥】の代わりに【￣】を追加し、それ以外は何も追加しなかった。

残り三つの表記方式は、ローマ字での表記を試みた。そのため、まずカタカナの発音形をローマ字表記へ変換する必要があった。

ローマ字表記には、日本語学習者が親しみやすい改訂ヘボン式 (BGN&BCGN 2015) を応用した。ヘボン式では長音を曲折アクセント記号（＾）か、長音記号（マクロン、￣）で記すが、この表記では長音の間に高低差がある場合、表示しにくいので、同じローマ字を二つ続けて書くことで長音を表した。

まずインプット用にカタカナ（大文字）からローマ字への変換表と、大文字と「ッ」以外の小文字との組み合わせからローマ字への変換表を用意した。次にそれらをそれぞれの Python の辞書（dictionary）へと変換した。次にアクセント表記付きのカタカナのリストの一単語ずつ繰り返し処理し、単語の一文字ずつ以下のように処理した。

処理している文字が特例に当たらない場合は処理している文字と次の文字が大文字と小文字のローマ字への変換辞書に入っている場合はそのように変換し、処理している文字が長音符か【ッ】の場合、そして【ン】に【アイウエオヤユヨ】のどれかが続く場合、もう一つの Python 辞書を使って変換した。長音符の場合、処理中の単語の変換済みの部分の最後のローマ字をそのままもう一つ足した。処理中の文字が【ッ】の場合、これが単語の最後の文字の場合、最後にビックリマークを追加した。例えば「アッ」は「a!」に変換した。【ッ】が最後の文字でない場合、次の文字が【シ】の場合【s】を追加し、【チ】の場合【t】を追加し、それ以外の場合は次の反復で辞書を使ってローマ字へ変換するときローマ字の一文字目を２文字にするようフラグを作った。最後の特例は、処理中の文字が【ン】の時にフラグを作り、次の反復で処理中の文字が【アイウエオヤユヨ】のどれかの場合は変換する前に【'】を追加した。これでローマ字へ変換された時にナ行とンの区別が付くようにした。

このように『日本語教育』の単語を上記のようにローマ字へ変換しながら、３種類のアクセント表記を追加した。

まず、最も実装しやすいであろうアクセント表記として、高く発音される拍をローマ字大文字で、低く発音される拍をローマ字小文字で表す表記方式を用意した。最初に作ったアクセント付きカタカナ表記の語彙リストをインプットとして使い、上記のローマ字への変換のアルゴリズムに【˥】と【˥】の記号を区切り符号として処理する仕組みを足し、最初の拍を変換する時、一番目の区切り符号が【˥】だった場合は小文字のままにし、

【⏋】だった場合は大文字に変えた。単語をローマ字へ変換するときの反復処理の各反復では前の反復で小文字へ変換していたか、【⏋】だった場合は小文字へ、【「】だった場合と大文字へ変換していた場合は大文字のローマ字へ変換した。

次に試したアクセント表記は、高く発音される拍のローマ字を高い位置に書き、低く発音される文字を低い位置に書くやり方である。できるだけ多くの環境で利用できる汎用性の高いデータを目指しているので、高音の文字を Unicode に含まれている上付きローマ字 (superscript) で表し、低音の文字は通常のローマ字で表す表記を用意した。低音の文字を Unicode の下付きローマ字 (subscript) で表す表記も試みたが、Unicode に登録されているローマ字下付き文字はごく一部でしか存在しない為、実装できなかった。幸い、上付きローマ字は q 以外全ての小文字があった為、上付きローマ字と通常の小文字ローマ字の組み合わせを使った。

まず大文字を Unicode の上付きローマ字小文字へ変換する Python の辞書を作った。次にインプットとして大文字と小文字を使ったアクセント表記の単語リストを使い、各単語を反復処理し、大文字を上付きの小文字へ変換した。上付き小文字が非常に小さいため、上付き大文字に変換したリストも用意した。ただし、上付き大文字の c,y,s が存在しないため、その文字を上付き小文字にした。

最後に試した表記方式は、高く発音される拍を別の色で表す方法である。インプットとしては上記と同じように大文字と小文字のローマ字表記のリストを作った。それぞれの単語の大文字の部分をタグで囲い、全てを小文字に変換した。次にリストを HTML の表に変換し、囲いのタグを文字を赤色に変えるタグに変えた。これで HTML ファイルを開きリストをコピーすれば、ほとんどのエディターへペーストできる。今回のデータは全て Google スプレッドシートにも保存したが、Google スプレッドシートは HTML が使えなかったため、最初に作ったタグがついたリストをスプレッドシートに書き込みスクリプトエディタ「Google Apps Script」を使いフォーマットを施した。

このように装備したデータの一部を、図 3 に示す。

標準的な表記	UniDic 発音形出現形	アクセン	UniDic 発音形出現形＋アクセント「⏋	UniDic 発音形出現形＋アクセント＼	大小ローマ字	上下付きローマ字	上下付き 小文字	色分けローマ字
かかわる	カカワル	3	カ「カワ⏋ル	カカワ＼ル	kaKAWAru	ka^{KAWA}ru	ka^{kawa}ru	kakawaru
柿	カキ	0	カ「キ	カキ	kaKI	ka^{KI}	ka$^{k^{i}}$	kaki
夏期	カキ	1	カ⏋キ	カ＼キ	KAki	^{KA}ki	^{ka}ki	kaki
火気	カキ	1	カ⏋キ	カ＼キ	KAki	^{KA}ki	^{ka}ki	kaki
夏季	カキ	1	カ⏋キ	カ＼キ	KAki	^{KA}ki	^{ka}ki	kaki
垣	カキ	2	カ「キ⏋	カキ＼	kaKI__	ka^{KI}__	ka$^{k^{i}}$__	kaki__
かぎ	カギ	0	カ「ギ	カギ	kaGI	ka^{GI}	ka$^{g^{i}}$	kagi
餓鬼	ガキ	0	ガ「キ	ガキ	gaKI	ga^{KI}	ga$^{k^{i}}$	gaki

図 3 アクセント表記入り単語リストの一部

４．考察

４.１　それぞれのアクセント表記方式の長所と短所

多くの教科書で使われている【「】と【﹁】の表記方式は、ほとんどの環境で使えてどのエディターでも簡単に入力でき、アクセントの上がり下がりも直感的にわかりやすい。短所としては一番最初の拍が高音か低音なのかがすぐにはわからないことと、フォントによっては【「】と【﹁】という記号が見えにくいことが考えられる。

『NHK 日本語発音アクセント新辞典』(2016)で採用されている【＼】の表記方式は文字が小さくても見えやすいが、日本語アクセントの法則をよく理解していない学習者には、読み解くのが困難だと考えられる。

ローマ字の大文字と小文字の長所は、どのエディターでも使えて汎用性が高く、文字が小さくても読めることだと言えるが、大文字小文字の区別はオンラインコミュニケーションでは強弱や強調を表すのによく使うので、高低アクセントとして認識しにくいと考えられる。

上付きローマ字を使って高音を表す表記方式は、高低の差が直感的に伝わると考えら、Unicode に含まれる文字なので汎用性が高いが、全体的文字のフォントを大きくしないととても読みにくいことが短所だと言える。また、Unicode には上付き大文字の c, s, y が存在しないため、上付き大文字表記のリストでは上付き小文字の c, s, y を使ったが、文字の高さが揃わず読みにくい。

色分けの表記方式はフォントサイズが小さくても読めて、印象的だと考えられるが、別のエディターや環境へ移植しにくい、ポータビリティが低いことと、どの色が高音でどの色が低音かが直感的にわからないため、記憶、認識しにくいと思われる。

４.２　日本語学習者の反応

３で述べた過程で用意した語彙リストを、10 人の日本語学習者に提示し、どの表記が使いやすいか聞いてみたが、全くの初心者や文字が苦手な学習者はローマ字の上付き表記方式が読みやすいと答えた。上級者はローマ字の上付き表記と、鉤括弧（【「】と【﹁】）を使った表記方式を選んだ。色分けの表記方式を読みやすいと評価した学習者もいたが、数字でアクセントを表す表記方式を選ぶ人は 1 人もいなかった。

５．今後の課題

本稿では、日本語学習者用の辞書に掲載できるアクセント表記入りの日本語教育語彙データとその装備の過程を紹介し、５つのパイロット版のデータの使いやすさに対する数人の学習者から得た反応について報告したが、それぞれのアクセント表記データの利用効果、アクセント学習への影響を検証するためには、さらなる実験が必要である。

また、JEV のアクセントデータには誤字がいくつか発見され、そのデータの確認と、アクセント型が複数定着している単語の表示のありかたが今後の課題である。

謝　辞

本研究では，日本語学習辞書支援グループ (2021)「日本語教育語彙表 Ver 2.8.3」と、「現代語用の UniDic」を使った「Web 茶まめ」を利用した。データの整理に協力してくださったカタリーナ・ゲールさんに感謝する。

Kristina Hmeljak Sangawa's research was financially supported by the Slovenian Research and Innovation Agency (ARIS) in the research core funding Asian Languages and Cultures (P6-0243).

文　献

石黒圭・吉甜・佐野彩子編(2024).『スマホは辞書になりうるか –日本語学習者の辞書引きの困難点と指導法–』明治書院

神田卓朗 (2003).「日本語アクセントの表記に関する考察 —三線式表記法—」『岐阜女子大学紀要』32 巻, pp. 51-58.

堤智昭・小木曽智信 (2023).「複数の UniDic 辞書による形態素解析支援ツール『Web 茶まめ』の実装と運用」, 情報処理学会論文誌 64:3, pp.749-757. DOI 10.20729/00225271

NHK 放送文化研究所編 (2016).『NHK 日本語発音アクセント新辞典』NHK 出版

フメリャク寒川, クリスティナ (2019).「日本語非母語話者の読解コーパス」から見える非漢字圏日本語学習者の辞書使用」『言語資源活用ワークショップ発表論文集』351-358.

フメリャク寒川, クリスティナ・一宮由布子・井田尚美・古賀美千留・守時なぎさ・柳玹娵 (2012).『Japonščina za začetnike 初級日本語１』リュブリャナ大学出版局

BGN & PCGN (United States Board on Geographic Names and Permanent Committee on Geographical Names for British Official Use) (2015). Romanization of Japanese kana: Modified Hepburn System BGN/PCGN 1976 Agreement, modified in 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/320105/Japanese_kana_Romanisation.pdf

EDRDG – Electronic Dictionary Research and Development Group (2024) JMdict-EDICT Dictionary Project. http://www.edrdg.org/wiki/index.php/JMdict-EDICT_Dictionary_Project

jaSlo – Kristina Hmeljak, Tomaž Erjavec, and Irena Srdanović, Irena (2016). Japanese-Slovene learner's dictionary jaSlo 3.1, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1050

JEV – 日本語学習辞書支援グループ (2021) 「日本語教育語彙表 v.2.8.3」 http://jhlee.sakura.ne.jp/JEV/

OJAD – オンライン日本語アクセント辞典 https://www.gavo.t.u-tokyo.ac.jp/ojad/

Yuriko Sunakawa, Jae-Ho Lee, and Mari Takahara (2012). “The construction of a database to support the compilation of Japanese learners' dictionaries”, Acta Linguistica Asiatica 2:2, pp. 97-115. https://doi.org/10.4312/ala.2.2.97-115