



日本語学習者のための「日本語音便データベース」 の作成

著者	池谷 知子, 黒木 邦彦, 田附 敏尚
著者別名	IKEYA Tomoko, KUROKI Kunihiko, TATSUKI Toshihisa
雑誌名	トークス = Theoretical and applied linguistics at Kobe Shoin : 神戸松蔭女子学院大学研究紀要言語科学研究所篇
巻	22
ページ	25-32
発行年	2019-03-05
URL	http://doi.org/10.14946/00002083

日本語学習者のための「日本語音便データベース」の作成*

池谷知子・黒木邦彦・田附敏尚

神戸松蔭女子学院大学 文学部

tikeya[at]shoin.ac.jp · kujonjaroo9215[at]shoin.ac.jp · tatsukitatu[at]shoin.ac.jp

Making of “the Japanese Sandhi Database” for Japanese Learners

IKEYA Tomoko · KUROKI Kunihiko · TATSUKI Toshihisa

Kobe Shoin Women’s University Faculty of Letters

Abstract

本稿では、日本語の日常会話に頻出する音便を集めたデータベース（以下「音便 DB」）の利用方法および有用性を紹介する。このデータベースは、漫画、アニメ、映画などをきっかけに日本語学習を始める学習者の使用を想定し、筆者たちが作成を進めているものである。「すげー ひやっと すっから（←すごい ひやっと するから）」のような音便は、非規範的な日本語とされ、従来の日本語教育ではあまり扱われることがなかった。そのため、学習者がこのような音便を含む台詞を理解しようとしても、「すげー」「すっから」を自分自身で調べる手立てがない。そこで、音便 DB は、一般的な国語辞典に拠れば、検索語の構成要素を掴みうる構成とした。同時に、検索精度向上と自立学習支援とを狙って、音便の音韻的・形態的条件も入力している。音便 DB は、ソフトの普及度やデータ入力の簡易性の面から、今のところ、MS Excel® に拠ったパイロット版である。将来的には、スマートフォン経由の潜在的利用者也取り込むため、学習者が web 上で気軽に検索できる形式で公開することを目指している。

In this paper, we introduce “the Japanese Sandhi Database” (hereafter “the Sandhi DB”), which contains many examples of sandhi often used in daily conversation, while explaining its usage and usefulness. This database, under production, intends to help L2 learners of Japanese who start studying Japanese through comics, animation and movies. Sandhi forms, such as *sugee hiyaQto suQkara* ‘because

*本稿は、2018 年 8 月 3 - 4 日にヴェネチア大学で行われた日本語教育国際研究大会（Venezia ICJLE 2018）におけるポスター発表を改訂したものである。

it's really chilly', are regarded as non-standard Japanese and hence are hardly covered in conventional Japanese L2 education. When learners cannot understand dialogue with the sandhi forms cited above, there is no way to consult a dictionary for *sugee* (← *sugoi* 'very') and *suQkara* (← *surukara* 'because I will do so'). For this reason, the Sandhi DB consists of data to assist users in understanding the constituents of search words according to common Japanese dictionaries. It also contains phonological and morphological conditions of sandhi for improving the search accuracy of the Sandhi DB and for supporting learners' autonomous learning. The database is currently in beta version created in Microsoft Excel® in order to easily share the database and simplify data entry; however, in the future, the authors aim to make the database available in a format which can be accessed via smartphones, so that L2 learners can easily access its content and search for Japanese sandhi forms on the web.

キーワード: 日本語教育、音韻論、形態音韻論、連声、自然音類

Key Words: Japanese language education, phonology, morphophonology, sandhi, natural class

1. 研究の背景

近年、海外でも日本の漫画、アニメ、映画の人気の高まるにつれ、それらをきっかけに日本語の勉強を始める学習者が増えている。しかし、その中で使われる日本語は規範的な日本語ではないために、学習の障害になるものも多く含まれている。その中の1つに音便がある。「取りて→取って」のようなテ形音便は規則的かつ頻繁に起こるため、日本語学習の過程で学ぶ。しかし、「するから→すっから」「すまない→すまねー」のような不規則に起きる音便の理解・習得は、学習者個人に委ねられている。

また、漫画などに含まれる「すっきゃねん」などのよく知られている方言形は、日本語母語話者ならば簡単に理解できるが、辞書の記述の対象外となっている。

このように、漫画などの創作物で使用される日本語は、日常会話でよく聞かれるものの、規則性があまり見られない省略・縮約形を多数含んでいる。それらは非規範的な日本語とされ、従来の日本語教育ではあまり扱われることがなかった。そのため、学習者そのような台詞を理解しようとしても、自分自身で調べる手立てがない。学習者が「すっから」「すまねー」のようなものを見つけ、それを調べようとしても、規範的な形ではないために調べられないというジレンマが存在するのである。

2. 本研究の目的

そのような状況を打破するため、筆者たちは、日本語学習者が漫画などで見つけた、分からない日本語を自主学習するためのデータベースを作成しようと考えている。このデータベースは、日本語学習者が辞書を引いても載っていない役割語や省略・縮約形などに含まれる音便について広く集め、それらの構造を分析的に示す。「すっから」が「する+から」に、「すまねー」が「すむ+ない」に分析できることが分かれば、簡単に理解できるか、あるいは、辞書で調べることが可能になると考える。更に、「ない」という形

式が「ねー」となりやすいことを習得すれば、「知らねー」「いらねー」のような表現を見聞きした場合に、学習者自身が音便形の構造や意味を調べることが可能になる。

そして、データベースを作るだけでなく、「日本語音便データベース」（以下「音便DB」）として公開することも目指す。本稿では、その β 版（以下「音便DB β 」）の作成の成果および課題を示す。それとともに、話し言葉で不規則に起こる音便の傾向を集めたデータから明らかにする。

3. 「日本語音便データベース β 版」について

音便DB β は、無作為に選んだ3冊の漫画（使用文献については論文末を参照）から抜き出した827件の音便をデータベース化したものである。ソフトの普及度やデータ入力の簡易性の面から、今のところ、表1のようにMS Excel[®]に拠っている。

表1: 日本語音便データベース β 版 (MS Excel[®])

しら ことば 調べる言葉 Words you check		せつめい 説明 Explanations			おと か きそく 音が変わる規則 Rules for sandhi			
おと 音 (ローマ字)	ことば 文字 (ひらがな)	おと 音 (ローマ字)	いみ 意味	じしょがた 辞書形	まえ 前	おと 音1	おと 音2 (元)	あと 後
Sounds	Hiragana	Sounds	Meanings	Dictionary forms	Before	Variation	Standard	After
sumanee	すまねー	sumanai	済まない	すむ 自 ない 形	n	ee	ai	
suQkara	すっから	surukara	するから	する 他サ変 から 接助	su	Q	ru	kara
suNno	すんの	suruno	するの	する 他サ変 の 終助	su	N	ru	no

音便DB β は、【調べる言葉】【説明】【音が変わる規則】という3つのメインカテゴリーからなる。そして、これらの3つのメインカテゴリーは、以下のサブカテゴリーによって構成されている。

表2: 音便DB β のカテゴリー	
メインカテゴリー	サブカテゴリー
調べる言葉	音 (ローマ字), 文字 (ひらがな)
説明	音 (ローマ字), 意味, 辞書形
音が変わる規則	前, 音1, 音2 (元), 後

各カテゴリーの下位分類は次のように定義される。

1. 【調べる言葉 words you check】

- 音 (ローマ字) sounds
- 文字 (ひらがな) hiragana
当該語の表層形を音素表記ないしかな表記したもの。

2. 【説明 explanations】

- 意味 meanings
直訳を意識した当該語の標準語訳。

- 辞書形 dictionary forms
当該語の構成要素を品詞分解したもの。

3. 【音が変わる規則 rules for sandhi】

- 前 before
音便箇所直接先行する音素ないし音節の基底形。音便箇所が介音・音節主音 (medial, nucleus) であれば、頭音 (initial) に当たる音素 (e.g. su.g-oi :: su.g-ee ‘凄い’; u.t-eba :: u.c-yaa ‘打てば’) を、そうでなければ、音節を入力。
- 音 1 variations
音便箇所の表層形 (実際に見られる音便形)。
- 音 2 (元) standards
音便箇所の基底形 (規範的な形)。
- 後 after
音便箇所続く全音節の基底形。

なお、【調べる言葉】【音が変わる規則】におけるローマ字は、以下の表記法による音素記号である。

表 3: 音素表記法

a1.	長母音の音素表記は同一母音字の連続 (e.g. ii, uu, ee, oo, aa)。
a2.	長母音のかな表記には「ー」を使う (e.g. いー、うー、えー、おー、あー)。
a3.	促音の表記には小さい「っ」ではなくQを使う。
b1.	チ、ツはそれぞれci、cu (cf. *ti、*tu)。
b2.	チャ、チュ、チヨはそれぞれcya、cyu、cyo (cf. *tya、*tyu、*tyo)。
c.	助詞「は」「を」「へ」=「wa」「o」「e」(cf. *ha、*wo、*he)
d.	文末(と思われる箇所)を示す記号として を次のセルに打つ。

学習者がローマ字で「sumanee」を、あるいは、ひらがなで「すまねー」を検索すると、該当する【調べる言葉】が得られる。そして、その基底形が「sumanai すまない」であり、一般的な国語辞典に拠れば、「すむ (動詞)」と「ない (助動詞)」とに分析されることを【説明】から掴みうる。その音便の条件は【音が変わる規則】に記されている。¹たとえば、nai → nee という音便であれば、音便箇所の直前の音素 n を「前」に、音便箇所の表層形 ee を「音 1」に、音便箇所の基底形 ai を「音 2 (元)」に、音便箇所の直後の音素 (sumanee の場合は該当する音素がないので「0」) を「後」に示している。これらはいずれも音便の音韻的・形態的条件であり、音便 DB 完成版の検索精度向上のために入力している。

¹『旺文社 国語辞典』(第 10/11 版)の記述に従い、当該語の構成要素を次のように入力。[見出し語 同辞書における品詞] (e.g. [書く 他五],[て 接助],[な 終助])

表 4: 10 例以上見られた音便とその実例

音 2(元)	音 1	出現数	例
ai	→ ee	375	知らない > しらねー
ru	→ N	66	するな > すんな
no	→ N	43	ほしいもの > ほしいもん
ru	→ Ø	35	どうするんだよ > どうすんだよ
Ø	→ Q	28	いけない > いっけない
oo	→ o	27	だるう > だる
oiu	→ uu	19	というか > つーか
ae	→ ee	15	おまえ > おめー
ewa	→ ya	15	それは > そりゃ
oi	→ ee	15	すごい > すげー
ra	→ N	14	わからない > わかんねー
ru	→ Q	11	うるさい > うっせー

4. 分析の結果

この研究の1つの成果として、漫画などの話し言葉に散見される音便が実際にどのような形で現れているかを観察できる点が挙げられる。どのようなルールが存在するかの記述は先行研究にもいくつか見られるが（小磯・西川・間淵 2006、宮崎・佐藤 2018）、量的な調査結果は示されていない。量的な調査結果は、学習者に優先的に教えるものを選ぶにあたって、一つの指標となるはずである。

4.1. 出現数の多い音便

827 件のデータのうち、10 例以上見られたものをまとめると、【表 4】(p. 29) のようになる。「知らない (ai) → しらねー (ee)」のように、連母音 ai が長母音 ee に変わるパターンが最も多く、375 件 (45%) 見られた（個々の実例は次の 4.2 節で見ると）。この変化自体は珍しいものではないが、規範的な書き言葉に現れることはあまりない。今回用例を採取した資料が漫画であったため、このようなものも多く現れたと考えられる。

この他、連母音 ae が長母音 ee になる「おまえ (ae) → おめー (ee)」が 15 件、連母音 oi が長母音 ee になる「すごい (oi) → すげー (ee)」が 15 件あった。連母音 (ai, ae, oi) が長母音 (ee) になるパターン以外にも、「というか (oiu) → つーか (uu)」のように、oiu の音が uu に変化するものも 19 件見られた。これらを合計すると、連母音の長母音化は 488 件 (59%) となり、今回得られた音便の半数を超える。

よく見られる音便パターン①: 連母音の長母音化

ai, ae, oi → ee

oiu → uu

表 5: ai → ee の形態素別出現数

辞書形	品詞	出現数	辞書形	品詞	出現数	辞書形	品詞	出現数
ない	形	152	うまい	形	2	じゃまくさい	形	1
ない	助動	147	でかい	形	2	くだらない	形	1
うるさい	形	27	めんどくさい	形	2	おっかない	形	1
たい	形	7	いたい	形	2	だらしない	形	1
しかたない	形	6	うざい	形	2	なさない	形	1
みたい	接尾	6	うざったい	形	1	もったいない	形	1
ぜったい	副	4	どんくさい	形	1	きたない	形	1
あぶない	形	3	つまらない	形	1			
やばい	形	2	つらい	形	1			

次に多かったものは、「する (ru) な→すん (N) な」のように、n の直前で ru が撥音 N に変わるパターンで、66 件 (8%) 見られた。同じようなパターン (rV → N/ __n) として、「わから (ra) ない→わかん (N) ねー」も 14 件 (2%) 見られた。これらとは異なり、名詞「もの」において頻発する「ほしいもの (no) →ほしいもん (N)」などの 43 件 (5%) も合計すると、撥音化は 123 件 (15%) となる。

よく見られる音便パターン②: 撥音化

ru, ra → N/ __n
no → N

ru においては、撥音化だけでなく、「どうする (ru) んだよ」「どーす (Ø) んだよ」のような脱落も 35 件 (4%) 見られた。ru は音便化しやすく、撥音化・脱落に加えて、「うる (ru) さい→うっ (Q) せー」のような促音化も 11 件 (1%) 観察された。

よく見られる音便パターン③: ru の音便

ru → N, Q, Ø

上記のもの以外に見られた音便として、「い (Ø) けない→いっ (Q) けない」のような促音挿入が 28 件 (3%)、「だろう (oo) →だろ (o)」のような短母音化が 27 件 (3%)、「それは (ewa) →そりゃ (ya)」のような縮約も 15 件 (2%) 見られた。

4.2. ai → ee で出現数が多かった語

4.1 で取り上げた「ai → ee」という音便は、本調査最多の 375 例に上り、全体の 45% を占める。この音便がどのような語で見られたかを整理すると、表 5 のとおり、形態素別出現数にかなりの偏りがあると分かった。

まず、最も多い例は、形容詞「ない」が「ねー」になるものである。375 件のうち 152 件 (40%) がこの例であった。次に多い例は、助動詞「ない」が「ねー」に（「行かない」

が「行かねー」になるもので、147例(39%)見られた。つまり、否定の意味を持つ「ない」の音便が全体の79%を占めている。このことから、初級の学習者に対しては、話し言葉における否定の「ない」から「ねー」への音便を、複雑な音便規則には触れず、パターンとして教えることも可能である。

5. 今後の研究

前述のとおり、音便DB^βはソフトの普及度やデータ入力簡易性の面から、今のところ、MS Excel[®]に拠ったパイロット版である。以下、今後の研究の方向性を、「1. データの拡充」「2. オンライン化」「3. 音便の語彙的条件の解明」という3つの点から述べていく。

まず、データの拡充について述べる。今回は無作為に選んだ3冊の漫画から用例を集めた。今後、様々なジャンルに調査を広げると同時に、方言、男性語、女性語、役割語といった様々な位相語を積極的に取り込み、音便DBにその情報を反映させようと考えている。データ量を増やすことで、様々な変化形に対応でき、検索の質の向上が期待できる。

次に、音便DBのオンライン化について述べる。現在は、PC上での使用に留まっているが、スマートフォン経由の潜在的利用者も取り込むため、オンライン辞書のように学習者がweb上で気軽に検索できる形式で公開することを目指している。その際には、たとえば、ある語の音便を表示するときには、同種の音便—具体的には【音が変わる規則】の音1と音2(元)で共通するもの—を類例として表示するなど、検索結果の表示に、より工夫を凝らすことも必要であろう。このような機能は、学習者の自律学習に大きく寄与するであろう。

最後に、音便の語彙的条件の解明について述べる。標準語における一般的な音便として、イ音便(急ぎて→急いで)、撥音便(読みて→読んで)、促音便(言ひて→言って)、ウ音便(問ひて→問うて)が知られている。これらの出現条件は先行研究において解明されている。また、こうした規則的な音便は、正書法に基づく規範的な書き言葉にも見られる。

しかし、今回扱った音便の中には、規則的・規範的な音便とは異なり、イレギュラーに起こるものや、話し言葉にのみ現れるもの散見する。これらは従来、非規範的なものと見なされ、日本語教育の場では扱われなかった。今後は、こうした不規則的・非規範的な音便も含めて、どのような音素・形態素(連鎖)で音便が起こるのかを解明していきたい。

参考文献

- 小磯 花絵・西川 賢哉・間淵 洋子(2006)「転記テキスト」『日本語話し言葉コーパスの構築法』(国立国語研究所報告書 No. 124: 第2章)国立国語研究所, pp. 23–132.
斎藤 純男(1986)「話し言葉におけるラ行音およびナ行音のモーラ音素化」『日本語教育』60, pp. 205–220.

宮崎 千明・佐藤 理史 (2018) 「テキストのキャラクター性付与のための音変化現象の分類」
『言語処理学会 第 24 回年次大会発表論文集』 pp. 1227–1230.

Authors' web sites: <https://researchmap.jp/read0064389/>
<https://researchmap.jp/96ki/>
<https://researchmap.jp/7000003522>

(受付日: 2019 年 1 月 10 日)