

HaoriBricks3 による音変化表現の生成

佐藤 理史 柳 将吾 夏目 和子

名古屋大学大学院工学研究科

ssato@nuee.nagoya-u.ac.jp

1 はじめに

小説や漫画の発話テキストには、話し言葉固有の音変化が文字として現れる。宮崎ら [1] は、これを音変化表現と呼んでいる。実際に、小説の会話を調べると、多くの実例が観察される。以下に、『舟を編む』 [2] に見られる実例を示す。

- (1) 辞書はイメージもいいし、景気にも左右されにくい商品だってのに。(p12)
- (2) 『大渡海』も、後半がすかすかになっちゃうんですか？(p191)
- (3) はじめての辞書づくりにしては、岸辺君はいい線いっとるぞ。(p194)

このような音変化表現の生成は、小説の自動生成 [3] を実現するためには不可欠である。そこで、音変化表現を HaoriBricks3 (HB3) [4] で生成する方法を定め、宮崎ら [1] が示した音変化表現の 127 パターンのうち、120 パターンを生成できるようにしたので報告する。

2 基本方針

音変化表現には、変化前の表現が存在すると仮定し、音変化表現を変化前の表現の異形とみなす。この考え方に基づけば、元の表現から異形を作る操作を規定すればよいということになる。

しかしながら、音変化表現には、次のようなバリエーションが存在する。

- 内容語に関わるものと機能語に関わるもの
- 使用頻度が高いものと低いもの
- 単純なもの複雑なもの

このため、すべての音変化表現を統一的に扱うのではなく、複数の方法で扱うこととする。

具体的には、次に示す 5 つの方法で扱い、それぞれの方法を適用する原則を定める。

1. (新たな) 内容語として扱う

- 活用しない内容語が音変化を受けている場合
- 活用する内容語の語幹が音変化を受けている場合

2. 音変化形 (異形) を作り出すコマンドで扱う

- 音変化の形式が典型性を持つ場合

3. 新たな活用形として扱う

- 活用語尾が後続する語と縮約する場合
- 特定の活用形に限り、語幹の末尾の母音が音変化を受ける場合

4. 新たな機能語として扱う

- 機能語が音変化を受けている場合

5. テンス有標 (タ) の拡張として扱う

- 「～たあ」、「～たあ」

以下では、5 つのそれぞれの方法と保留としたパターンについて説明する。なお、以下において、「P_n」という表記は、宮崎らのパターン番号を表し、* 付きの場合は、生産性のあるパターンを示す。

3 新たな内容語

HB3 において、内容語の表層文字列を生成するために必要な情報は、以下のとおりである。

- 活用しない内容語：表記
- 活用する内容語：(辞書形) 表記、および、活用型

HB3 には、内容語の辞書形表記から活用型を自動推定する機能が組み込まれているため、辞書形表記さえあれば、表層文字列を生成することができる。HB3 のブリックコードでは、内容語を辞書形表記で記述するので、HB3 システムに何の変更を加えることなく、新たな内容語を扱うことができる。

たとえば、「すっごい (P10)」は「すっごい」という内容語として扱う。これを述語とみなす場合、活用型は「イ形容詞型アウオ段」と推定されるので、連用形「すっごく」などを正しく生成できる。

なお、「すっごい」は、次に説明する音変化ブリッ

表1 音変化ブリックの一覧

分類	ブリック	操作	宮崎ら [1]
促音化	イ音促音化	一番後ろのイ音を「っ」に置き換える	P55, P56
	ウ音促音化	末尾の「う」を「っ」に置き換える	P54*
	ル音促音化	末尾の「る」を「っ」に置き換える	P53*
	指示詞促音化	指示詞の一部を「っ」に置き換える	P57, P58, P59
撥音化	ニ撥音化	末尾の「に」を「ん」に置き換える	P73*
	ラ撥音化	末尾の「ら」を「ん」に置き換える	P82*
	ル撥音化	末尾の「る」を「ん」に置き換える	P86*
	指示詞撥音化	指示詞の一部を「ん」に置き換える	P76, P84
長音化	ア長音化	末尾に「あ」を追加する	「食べりゃあ」
脱落	イ脱落	末尾の「い」を削除する	P18*
	ウ脱落	末尾の「う」を削除する	P24*, P25, P26, P27, P28
	ク脱落	末尾の「く」を削除する	P19*
	ル脱落	末尾の「る」を削除する	P29*
	先頭文字脱落	先頭の文字を削除する	P38, P41, P42
促音挿入	チ促音挿入	一番後ろの「ち」の直前に「っ」を挿入する	P12*
	テ促音挿入	一番後ろの「て」の直前に「っ」を挿入する	P12*
	促音付加	末尾に「っ」を追加する	P9*
	頭2 促音挿入	先頭の文字の直後に「っ」を追加する	P10*
撥音挿入	頭3 促音挿入	先頭から2文字目の直後に「っ」を追加する	P11*
	頭2 撥音挿入	先頭の文字の直後に「ん」を追加する	P15, P17
記号挿入	末2 長音記号挿入	末尾の文字の直前に「ー」を挿入する	P2*, P3, P4, P7, P8
	末3 長音記号挿入	末尾の2文字目の直前に「ー」を挿入する	P5*
母音交替	AE交替	AIをEEに変更する	P43*
	OE交替	OIをEEに変更する	P45*
	UI交替	UIをIIに変更する	P44*
	セシ交替	「せ」を「し」に変更する	P46, P47
子音交替	サチャ行交替	サ行がチャ行に変更する	P48*
	ツチュ交替	「つ」を「ちゅ」に変更する	P49*
	リシ交替	「り」を「し」に変更する	P50*

クを用いて、「すごい」の2文字目に促音「っ」を挿入すると記述してもよい。ただし、ブリックコードを書く時点で「すっごい」を生成することが決まっているのであれば、直接「すっごい」と記述する方が簡単である。

4 音変化形を作り出すコマンド

音変化は、音という実在の物理現象に基づくため、ある種の規則性に則っている場合が多い。たとえば、末尾の文字の直前に長音記号が挿入される（つまり、最後から2番目の音が長音化される）現象は、「でーす」「まーす」「ませーん」「ごめーん」のような例に共通して観察される。これらの現象は、個別に扱わずに、そのような異形を生成するコマンドとして一括して扱う。これを**音変化コマンド**と呼び、それをHB3のブリックとして実装したものを**音変化ブリック**と呼ぶ。表1に、実装した音変化ブリックの一覧を示す。

音変化ブリックに組み込まれた音変化コマンドは、string_trans という名前の属性として保持さ

表2 音変化ブリックの使用例

ブリックコード	生成文字列
:ぼっちり	ぼっちり
:ぼっちし	ぼっちし
リシ交替 (:ぼっちり)	ぼっちし
意志形 (述語 (:走る))	走ろう
ウ促音化 (意志形 (述語 (:走る)))	走ろっ
述語 (:うるさい)	うるさい
AE交替 (述語 (:うるさい))	うるせえ
AE交替 (連用形 (述語 (:うるさい)))	うるさく

れ、HB3の語の表層文字列化の最終段階で実行される（そのような機能を新たにHB3に組み込んだ）。より具体的には、

- 活用しない語は、表記（文字列）を書き換える。
- 活用する語は、活用形の表層文字列を生成した後、その文字列を書き換える。

音変化ブリックの使用例を表2に示す。この例のウ促音化では、「走る」から意志形「走ろう」が生成され、その後、末尾の「う」を「っ」に書き換える。AE交替は、終止形「うるさい」を「うるせえ」に書き換える。連用形を指定した場合は、連用形「う

表3 音変化に関わる活用形 (1)

大分類	中分類	活用形	母音動詞型	子音動詞型	イ形容詞型	判定詞型	宮崎ら [1]
文末表示	推量形	タ系推量形	食べたろう	書いたろう	寒かったろう	—	P40
節末表示	条件形	条件音便形	食べりゃ	書きゃ	寒けりゃ	—	P98*, P99*
		条件縮約音便形	—	—	寒きゃ	—	P97*
	連用形	タ系連用チャ形	食べちゃ	書いちゃ	寒くちゃ	—	P92*, P94*
		タ系連用ジャ形	—	—	—	—	P93*

表4 音変化に関わる活用形 (2) 新たな活用形

大分類	中分類	活用形	特殊型 ぬ	イ形容詞型	子音動詞型 ワ行	ワ行文語音便	宮崎ら [1]
文末表示	タ形	タウ音便形	—	—	しもうた	問うた	
節末表示	条件形	条件縮約形	な	—	—	—	P20*
		連用形	—	さむか(ない)	—	—	P117
	連用ウ音便形	—	さむう	しもう	問う		
	連用ウ音便テ形	—	さむうて	しもうて	問うて		
	連用ウ音便省略テ形	—	—	—	—		
		連用ウ音便タリ形	—	さむうたり	しもうたり	問うたり	P87-90* P127*

るさく」が生成され、これには、アエ交替が適用できないため、そのまま「うるさく」が生成される。

5 音変化に関わる活用形

音変化表現の中には、活用語尾に関わるものが存在する。活用する語の活用語尾が後続する語と縮約する場合、および、特定の活用形に限り語幹末尾の母音が音変化を受ける場合は、新たな活用形として扱うのが簡単である。表3と表4に、音変化に関わる活用形の一覧を示す。

表3は、HB3にすでに組み込まれていた活用形である。P40「食べたろう/食べたろう」は、「だろ」の「だ」が脱落したものと説明されているが、HB3ではタ系推量形とみなす。これは、イ形容詞型の「寒かろう」(推量形)、「寒かったろう」(タ系推量形)という形式との類似性を重視するからである。

HB3では、タ系連用チャ形、および、タ系連用ジャ形という活用形が設定されている。これは、活用形として扱うのがよいか、あるいは、「接続助詞ちゃ」「接続助詞じゃ」として扱うのがよいかは、判断に迷う。HB3の活用体系は、JUMANの活用体系を出発点としたので、現時点では活用形として残っているが、将来、変更する可能性はある。

表4は、今回、音変化表現を扱うために、新たに新設した活用形である。

条件縮約形は、「否定接尾辞ぬ」の活用型である「特殊型ぬ」にのみ設定する活用形で、次の例のように「ねば」が「な」に縮退した形式である。(宮崎ら [1]は、このパターン(P20)を「なければ」の「ければ」が脱落した形式と説明している。)

- (4) a. 食べねばあかん
- b. 食べなあかん
- c. 食べにゃあかん

なお、最後の「(食べ)にゃ」は、「否定接尾辞ぬ」の条件音便形として扱う。

連用縮約形は、「イ形容詞型接尾辞ない」を除くイ形容詞型の活用型に定義される活用形で、次の例のように、「くは」が「か」に縮退した形式である。

- (5) a. 寒くはない
- b. 寒かない

残りの活用形は、ウ音便に関わる活用形であり、その基本となるのが連用ウ音便形である。

イ形容詞型の辞書形は「～い」となるが、「い」の直前の(語幹末尾の)母音にはエ音を除く4種類があり、それぞれの連用ウ音便形は、次のようになる。

1. 「ア音+い」は「オ音+う」に交替する
2. 「イ音+い」は「ユ音+う」に交替する
3. 「ウ音+い」は「ウ音+う」に交替する
4. 「オ音+い」は、原則として、「オ音+う」に交替する。ただし、「オ音+おい」の場合は、例外的に「オ音+お」となり、末尾の「い」が脱落する。

一方、子音動詞型ワ行の辞書形は「～う」となるが、「う」の直前の(語幹末尾の)母音にはオ音を除く4種類があり、それぞれの連用ウ音便形は、次のようになる。

1. 「ア音+う」は「オ音+う」に交替する
2. 「いう」は変化しない(「イ音+う」は、「いう」しかない)

表5 連用ウ音便形

辞書形	連用ウ音便形
-AI わかい	-OU わこう
-II かなしい	-YUU かなしゅう
-UI さむい	-UU さむう
-OI かしこい	-OU かしこう
-OOI とおい	-OO とお
-AU わらう	-OU わろう
IU いう	IU いう
-UU くるう	-UU くるう
-EU うれう	-OU うろう

3. 「ウ音+う」は変化しない
4. 「エ音+い」は、「オ音+う」に交替する

以上をまとめたものを、表5に示す。

連用ウ音便テ形と連用ウ音便タリ形は、連用ウ音便形に「て」「たり」が付いた形式である。連用ウ音便形に「た」がついたタウ音便形は、子音動詞型ワ行およびワ行文語音便に設定する。連用ウ音便テ形の「う」が脱落した連用ウ音便省略テ形は、子音動詞型ワ行にのみ設定する。

6 音変化に関わる機能語

音変化に関わる機能語の一覧を表6に示す。以下に示すような「ったら」「っちゃ」の扱いは悩ましいが、接続助詞とみなすこととする。

- (6) a. 嫌だったら嫌だ
b. 嫌たら嫌だ
c. 食べるたら食べる
- (7) a. 嫌だっちゃ嫌だ
b. 嫌っちゃ嫌だ
c. 食べるっちゃ食べる

7 テンス有標の拡張

テンス有標を表す「～た」の形式が「～たあ」「～たぁ」となる場合は、それぞれ「タあ」「タぁ」という特別なブリックで指定する。これは、HB3のテンスの位置自動決定機構で、テンスが実体化される場所が定まることに起因する。このブリックを用いると、「タあ(ちゃう(述語(:食べる)))」は「食べちゃったあ」を生成するが、これに「丁寧」ブリックを付加すると、「食べちゃいましたあ」を生成する。

8 実装保留

宮崎らのリストのうち、実装を保留しているのは、次の7パターンである。

表6 音変化に関わる機能語・従属節

中分類	新設	機能語	元の形	宮崎ら [1]
大分類=助詞				
引用助詞	√	っと	と	P13
	√	て	って	P35
副助詞	√	にゃ	には	P81
接続助詞		けど		P39
	√	ったら	といったら	P114
	√	っちゃ	といえぼ	P111
連体助詞		って	という	P30
	√	ん	の	P75, P116
大分類=助動詞				
連体助動詞		んだ		P78
	√	とこだ	とこ(ろ)だ	P36
丁寧助動詞		っす	です	P70
テ助動詞		る	(い)る	P21
	√	く	(い)く	P22
	√	まう	(し)まう	P23
大分類=述語接尾辞				
ボイス接尾辞		れる	(ら)れる	P32
テ縮約接尾辞		じまう	(デ)しまう	P103
		じゃう	じまう	P104
		ちまう	(テ)しまう	P103
		ちャう	ちまう	P104
		とく	(テお)く	P102
	√†	どく	(デお)く	P102
	√	たげる	(テあ)げる	P106
	√	だげる	(デあ)げる	P106
	√	たる	(テや)る	P107
	√	だる	(デや)る	P107
	√	とる	(テお)る	P102
	√	どる	(デお)る	P102
	√	といで	(テお)いで	-
	√	どいで	(デお)いで	-
大分類=従属節				
補足節	√	ん	の	P77

† 中分類変更

P31, P34, P72, P112, P113, P120, P126

実装を保留した理由は、音変化としての典型性が低く、使用される頻度もかなり低いと判断したためである。

謝辞 本研究はJSPS科学研究費基盤研究(B)「日本語文章の構造モデルとその段階的詳細化による文章自動生成機構」(課題番号18H03285)の助成を受けている。

参考文献

- [1]宮崎千明, 佐藤理史. 発話テキストへのキャラクター性付与のための音変化表現の分類. 自然言語処理, Vol. 26, No. 2, pp. 407-440, 2019.
- [2]三浦しをん. 舟を編む. 光文社, 2011.
- [3]佐藤理史. コンピュータが小説を書く日. 日本経済新聞出版社, 2016.
- [4]佐藤理史. Haoribricks3: 日本語文を合成するためのドメイン特化言語. 自然言語処理, Vol. 27, No. 2, pp. 411-444, 2020.