

任意の日本語テキストを対象とした活用語アクセント辞書の自動生成 Automatic Generation of Accent Dictionary of Conjugational Words for Any Japanese Text

東京大学大学院	小林俊平	skobayashi@gavo.t.u-tokyo.ac.jp
東京大学大学院	清水信哉	s_shimizu@gavo.t.u-tokyo.ac.jp
東京大学大学院	鈴木雅之	suzuki@gavo.t.u-tokyo.ac.jp
東京大学大学院	峯松信明	mine@gavo.t.u-tokyo.ac.jp
東京大学大学院	広瀬啓吉	hirose@gavo.t.u-tokyo.ac.jp
吉林華橋外国語学院	平野宏子	hiran565@gmail.com

キーワード：日本語音声教育，教材開発，活用語，アクセント型

1. はじめに

近年の日本語発音教育の中心は単音の発音に関するものがほとんどであるが、より自然な日本語発音を獲得するためには、単語のアクセントや、その変形の様子を習得することも重要である。ピッチアクセント言語である日本語では、音の高さの変化パターンの様子が日本語らしさに大きく影響する。このようなアクセントやその変形を体系的に教育するためには、その教材開発も必要となる。筆者らはこれまで、特定の日本語教材に出現する活用語を対象に、活用によるアクセント変形を体系的に閲覧できる、オンラインアクセント辞書を主に人手で作成してきた。

一方で自然言語処理の分野では、日本語テキストを自動解析（形態素解析）するツールが広く用いられている。また音声工学の分野では任意の日本語テキストに対して、アクセント句境界やアクセント核位置を推定する技術が整備されつつある。

そこで本研究ではこれらの技術を用い、任意の日本語テキストに対して活用語（動詞，形容詞，形容動詞）を抽出し、各活用語が有する全活用パターンを、アクセント核位置を明示した上で表示するオンラインアクセント辞書を作成した。すなわち、従来人手で用意していたアクセント辞書を、任意の日本語テキスト中の活用語に対して対応できるよう、拡張を行った。

2. オンラインアクセント辞書

まずは筆者らが作成したオンラインアクセント辞書 OJAD(Online Japanese Accent Dictionary) [1]について説明する。従来のアクセント辞書では、そこに掲載されているのは辞書形のアクセント型のみで、各単語それぞれの活用形のアクセント型は、利用者が規則を当てはめて考える必要があった。それに対して OJAD の大きな特徴は、見出し語のアクセントに加えて、用言が活用した際のアクセント型も表示している点である（図 1 に OJAD の Web インタフェースを示す）。活用形の種類としては、日本語教育に用いられる代表的な活用形 10 個（～ます形，使役形，受動形など）を掲載している。これにより、教科書等に出現した用言に対し、その活用形のアクセント型を知ることができるだけでなく、その語が他の活用をした場合のアクセントも合わせて知ることができ、これは日本語学習者や日本語教師にとって非常に有用であると考えられる。しかし一方で、OJAD が扱っている語彙は特定の教科書（日本語能力試験の 4 級から 1 級までの出題範囲が主）であり、単語数が圧倒的に少なく、拡張性に乏しかった。

3. 日本語のアクセント解析に関する要素技術

本節では、任意のテキストの活用語アクセント辞書を作成するに当たり、使用する工学的な要素技術について説明する。

見出し	辞書	ます	て	た	ない	なかった	ば	使役	使身	可能	意志
会う・会います	あう	あいます	あって	あった	あわない	あわなかった	あえば	あわせる	あわる	あえる	あおう
合う・合います	あう	あいます	あって	あった	あわない	あわなかった	あえば	あわせる	あわる		
運う・運います	あう	あいます	あって	あった	あわない	あわなかった	あえば	あわせる	あわる		
願ぐ・願ぎます	あおぐ	あおぎます	あおいで	あおいだ	あおがない	あおがなかった	あおげば	あおがせる	あおがれる	あおげる	あおごう
仰ぐ・仰ぎます	あおぐ	あおぎます	あおいで	あおいだ	あおがない	あおがなかった	あおげば	あおがせる	あおがれる	あおげる	あおごう

図1 オンラインアクセント辞書 OJAD のスクリーンショット

3.1 形態素解析

日本語は英語のように単語同士が空白で区切られていないため、日本語の解析を行うにあたって、まずは文章を単語に分割する形態素解析が必要である。形態素解析に関しては、フリーで提供されている日本語辞書[2]と解析器[3]を用いることで、高い精度を達成している。

3.2 単語接続にともなうアクセント位置推定

単語と単語が結合すると、各単語は、単語固有の型とは異なるアクセント型を持つことが頻繁に起こる。本研究においても、活用語が活用する際に自立語と付属語の単語接続が起こるため、接続後のアクセント位置を適切に推定する必要がある。

アクセント位置推定に関する最も代表的な既存手法としては、匂坂ら[4]によるアクセント結合規則を用いた手法がある。匂坂らはテキスト音声合成を念頭において、日本語におけるアクセント結合の現象を網羅的に説明可能な規則を構築した。この手法は高い精度を示し、テキスト音声合成の分野で広く使われている。更に、この規則に対して機械学習を応用し、アクセント変形予測の更なる高精度化を検討している研究例もある[5]。

4. 活用語アクセント辞書の自動生成

3節で紹介した要素技術を用い、任意の日本語テキストに対して、それに含まれる全活用語を対象としたアクセント辞書を自動作成する。その手順は以下のとおりである。

- 1) 入力されたテキストに形態素解析を施し、活用語(動詞、形容詞、形容動詞)を全て抽出する。
- 2) 抽出された各用言の各活用形のアクセント型を、3.2のアクセント推定技術を用いて推定する。
- 3) 抽出された各用言を、各活用形のアクセント型と共に図1のように列挙した形で出力する。

これらの処理をWebアプリケーション上で統一的に実装した。これにより、利用者はテキストファイルやテキストのコピーを入力するだけで、そのテキスト中に出現した全ての活用語の活用形とそのアクセント型を知ることができる。

5. 今後の課題

今後の課題としては、利用者(教師及び学習者)に実際にアプリケーションを使用してもらうことにより、インタフェースを改善していく予定である。

参考文献

- [1] オンラインアクセント辞書 OJAD <http://maria.gavo.t.u-tokyo.ac.jp/modules/ojad/>
- [2] 形態素解析辞書 Unidic <http://www.tokuteicorpus.jp/dist/>
- [3] 形態素解析器 和布蕪(Mecab) <http://mecab.sourceforge.net/>
- [4] 匂坂ら：“日本語単語連鎖のアクセント規則”，電気通信学会論文誌 J66-D.7, pp. 847-856(1983)
- [5] 黒岩ら：“単独ラベラによる大規模アクセントデータベースの構築およびそれを用いた統計的アクセント結合処理の検討”，電子情報通信学会音声研究会, SP2006-174, pp. 31-36 (2007)