

語のつながりのアクセント規則を用いた

4D-5

仮名文字文の形態素解析*

重泉 有希 鈴木 由理 山口 昌也 乾 伸雄 野瀬 隆 小谷 善行 西村 恕彦
 (東京農工大学 工学部 電子情報工学科)

1 はじめに

自然言語処理において、処理する文が音声文である場合、重要な役割を果たすのがアクセント情報である。音声文には文字種などの表記情報がないため、形態素解析の精度が低くなることは避けられない。そこで、音声文をイメージした仮名文字だけからなる日本語文の形態素解析に、アクセント情報を利用することを試みた。

2 アクセントの性質

日本語アクセントにはアクセント型があり、大きく平板式と起伏式に分かれる。平板式は平板型(桜:サクラ)、起伏式はさらに頭高型(命:イノチ)、中高型(心:ココロ)、尾高型(雲:シズク)に分かれる [1]。

標準アクセントには、次の性質がある。

単語アクセント性質 一つの単語の中で、高く発音される部分が離れて二箇所以上現れない。

文節の切れ目性質 文節の切れ目が分かる高低の組合せがある。

例:キレイナサクラ(きれいな桜)

1 0 0 0 0 1 1 → 1 0 0 0 · 0 1 1

1:高い拍 0:低い拍

アクセントは個々の単語に決まっているが、語と語がつながり、文になるとそのアクセントは変化する。変化の仕方は、品詞やその語に決まっているアクセントの型など、語の種類による規則に基づいている。この規則を“接続アクセント規則”と呼ぶ。

*Morphological Analysis of Sentences consist of Only Kana Characters using Accent Rules about Connection between Words,

Yuki SHIGEIZUMI, Yuri SUZUKI, Masaya YAMAGUCHI, Nobuo INUI, Takashi NOSE, Yoshiyuki KOTANI, Hirohiko NISIMURA,
 Tokyo University of Agric. and Tech., Dept. of Computer Science

3 接続アクセント規則

語のつながりのアクセント規則は、つながる語の品詞、アクセントの型、拍数によって決められている [1]。本研究では、文献 [1] をもとに“接続アクセント規則”を規則化した。

「前部+後部」のそれぞれの品詞によって、“接続アクセント規則”が存在する。規則の例を次に示す。

接続アクセント規則 名詞+名詞

- 後部の名詞が中高型 →
後部の高さの切れ目まで高い
- 後部の名詞が中高型以外 →
後部の一拍目まで高い

「卒業」+「論文」をこの規則にあてはめると、

「ソツギョウ(平板)」+「ロンブン(平板)」
→「ソツギョウロンブン」

となる。また、後部が助詞や助動詞のときは、前部の品詞、アクセント型(平板と起伏の区別)と、後部の助詞の種類によって規則が存在する。

接続アクセント規則 名詞+助詞

例:助詞「が」の場合

- 前部の名詞が平板式 →
前部のアクセントに続き助詞が高くつく
- 前部の名詞が起伏式 →
前部のアクセントを生かし助詞が低くつく

「私が」と「僕が」をこの規則にあてはめると、

「ワタシ(平板)」+「ガ」→「ワタシガ」
「ボク(起伏)」+「ガ」→「ボクガ」

のようになる。

4 規則を用いた形態素解析

仮名文字文とそのアクセント(高い拍を1、低い拍を0と表す)を入力とする。仮名文字文を既存の形態素解析システムにより、品詞の接続から可能性のある形態素候補を得る。辞書は形態素辞書にアクセント情報を付加したものを使用する。アクセント情報の形式は品詞によって異なり、それぞれ次のようにした。

- 自立語(名詞、動詞、…)
その単語のアクセントの変化の位置
- 助詞、助動詞
“接続アクセント規則”で用いる種類

得られた形態素候補に、アクセント情報を利用した次の三段階の処理を行い、形態素解析候補を絞る。

形態素のアクセント 第一段階の処理は、“単語アクセント性質”を用いる。候補として切り出された形態素の入力アクセントが、単語のアクセントとして成り立つか調べ、候補を絞る。

接続アクセント規則適用 第二段階の処理は、“接続アクセント規則”を用いる。前部・後部、それぞれの単語の拍数やアクセント型から、「前部+後部」全体のアクセントを生成する。

アクセントの一致 第三段階の処理は、生成アクセントと入力アクセントが一致するか調べる。“文節の切れ目性質”から得られる入力文の文節の切れ目を参照し、形態素候補が文節の切れ目まで解析されているか調べる。もし、切れ目まで解析されていたら、第二段階の生成アクセントと入力アクセントが一致するか調べる。文節の切れ目のパターンは、複数得られるときもある。複数得られたときは、すべての切れ目のパターンにおいて処理をする。

アクセント規則適用例を次に示す。

解析入力文:イシガタイ(意思が固い)

入力アクセント:100010

解析候補中の候補1、2について取り上げる。

——— 文頭+イシ ———

候補1 「イシ(石)」、名詞、尾高型

候補2 「イシ(意思)」、名詞、頭高型

第一段階 候補1、2とも単語のアクセントとして成り立っている。文頭の語なので第二、第三段階は行わない。

解析候補中の候補1-1、2-1について取り上げる。

——— イシ+ガ ———

候補1-1 前部:「イシ(石)」、後部:「ガ」、助詞、アクセント情報[1,1,3](名詞、動詞、形容詞に接続するときの助詞の分類)

候補2-1 前部:「イシ(意思)」、後部:「ガ」、助詞、アクセント情報[1,1,3]

第一段階 候補1-1「ガ」、2-1「ガ」とも単語のアクセントとして成り立っている。

第二段階 名詞+助詞の“接続アクセント規則”を用いて、候補1-1「イシガ」のアクセント「中高型」を決定する。候補2-1「イシガ」のアクセント「頭高型」を決定する。

第三段階 “文節の切れ目性質”から得られた文節の切れ目(三拍目か四拍目で切れる)を参照し、候補1-1、2-1の解析はともに文節の切れ目(三拍目)まで終了しているので、アクセントが一致するか調べる。

候補1-1「イシ(石)ガ」

生成アクセント:010

入力アクセント:1000→失敗

候補2-1「イシ(意志)ガ」

生成アクセント:100

入力アクセント:1000→成功

5 まとめ

本稿では、アクセント情報を利用した仮名文字文の形態素解析における、語のつながりのアクセント規則の適用方法について説明した。本研究は科研費(一般研究C)の支援を受けて行われた。

参考文献

- [1] 金田一春彦ほか:明解日本語アクセント事典,三省堂,1994
- [2] 水野孝司:日本語アクセント生成システム,情報処理学会第40回全国大会講演論文集(I),pp.317-318,情報処理学会