

言語情報を考慮したアクセント型推定手法

蜂谷 雅弘* 森山 高明* 小川 均* 天白 成一** 橋本 雅行**

*立命館大学理工学部情報学科

〒525-77 滋賀県草津市野路東 1-1-1

** (株) アルカディア

〒562 大阪府箕面市坊島 1-3-40

日本語では、単語のアクセント型は、単語を構成する漢字やその読みの情報からある程度推定することができる。これは単語を構成する漢字が単語のアクセント型に影響しているためであると考えられる。一方でアクセント辞典には沢山のアクセントの法則が明らかにされている。これらアクセントの法則は言語的な情報や音韻的な情報を基に、単語アクセント型の傾向を決定する。本報告では構成漢字が単語に与えるアクセント型の影響や、これらのアクセントの法則によって決定されるアクセント型の傾向を考慮したアクセント推定手法を提案した。そしてその手法によって3モーラの一般名詞に対してアクセント型の推定を行ない90.3%の正解率を得た。

Accent type estimation method in consideration of linguistic information

Masahiro HACHIYA* Takaaki MORIYAMA* Seiichi TENPAKU** Masayuki HASHIMOTO**
Hitoshi OGAWA*

*Ritsumeikan University

1-1-1, Nojihigashi, Kusatsu, Shiga, 525-77 JAPAN

**Arcadia, Inc.

1-3-40, Bohnoshima, Minoh, Osaka, 562 JAPAN

Accent types of Japanese words are estimated from their constituent Chinese characters to a certain degree. On the other hand, there are many accent rules in Japanese accent dictionaries. A tendency of accent type of words is determined from linguistic information or phonetic information of words. An accent type estimation method was proposed, in consideration of these rules and the constituent Chinese characters in this report. And we reported that we can estimate the correct accent type for 3-mora Japanese common noun 90.3% percent by this method.

1 はじめに

広辞苑 [5] では「アクセント」は、「特定の音節の特に際立った高まりや強まり」と定義されており、それぞれの語について定まっているとされている。これはどのような条件下でも完全に固定されることを意味しているわけではなく、地域や時代の移り変わりによって変化し、さまざまな方言を形成している。アクセントは単語の意味を区別したり、伝える相手

に言葉の理解を促す上で極めて重要で、人はこれらを頭の中でパターン(アクセント型)として整理し記憶している。[6] しかし人が全ての単語のアクセントを記憶しているわけでないことは明らかである。では記憶していない単語のアクセントはどのようにして得られるのであろうか?

文献 [3][4] には、単語のアクセント以外に多数のアクセントの法則が記載されている。アクセントの法

則とは単語についての何らかの情報を基にその単語のアクセントにどのアクセント型が多いなどの傾向をもたせる法則である。法則は方言によって様々であるが、同一の方言ではほぼ一定の法則となっている。したがって新しい語ができて、同じ方言の話者であればほぼ同じアクセントで発話される。アクセントの法則はその基になる情報によって以下のように大別することができる。

- 音韻的な情報を基にした法則
(単語の読みなど)
- 言語的な情報を基にした法則
(単語の品詞, 活用, 素性, あるいは漢語, 和語, 外来語といった語の戸籍など)

人はこれらの法則と基本となるいくつかの単語のアクセントを記憶し、その他の記憶していない単語については、これらから類推していると考えられている。

これまでのアクセント推定に関する研究では、日本人の姓を対象としたものがいくつか報告されている。1つは広川ら [1] によるものである。アクセント型と姓に用いられる漢字の間には相関があるものがあることから、姓に用いられる漢字の内頻度の高いものについて、アクセント型の傾向を定量的に明らかにした。出現頻度を考慮して語頭、語尾漢字のアクセント型傾向度合を求め、これに基づいて平板型か起伏型かの推定を行い、90.9%の正解率を得ている。もう1つは北川ら [2] の研究で、語頭漢字、語尾漢字の種類および文字位置で決まるアクセント型を付与する規則(一般則)と、頻出姓の平板化あるいは頻出地名の影響と思われる特なアクセント型を付与する規則(特殊則)を順次適用することによりアクセント型推定を行なう。以上の手法に基づき推定を行なった結果92.6%の正解率で正しいアクセント型が推定できたことが報告されている。

またこれまでの我々の研究では、参考に3, 4モーラの一般名詞を対象として研究を行なった。推定の手法として広川ら、北川らの手法を参考に、単語を構成する漢字やその読みに着目して作成した統計データより語頭漢字、語尾漢字のn型アクセント率を計算し、それらの積が最大となる型を推定型とする手法を用いた。その結果、3モーラ語で90.1%、4モーラ語で92.1%の正解率を得ることができた。

これらの研究では単語のモーラ数、単語の読み、各漢字の読み、語頭、語尾漢字といった情報からアクセ

ント型の傾向を求める手法によってある程度の成果を得ている。一方、アクセント辞典にはこれまでの手法では未だ利用していない音韻的な情報や言語的な情報を基にしたアクセントの法則が多く記載されている。これらの法則を推定手法にうまく取り入れることにより、より良い推定が可能となると考えられる。

そこで本報告では、これまでの推定手法を基本とした上で、様々なアクセント法則を取り入れることのできるアクセント推定手法を提案し、アクセント型の推定を行なった。

2 研究対象

本研究では、既にアクセント辞書に登録されている一般名詞のうち、以下のような条件を満たす語を分析および、アクセント型推定の対象とした。

- 3 または 4 モーラである。
- 漢字 2 字で構成される

その理由としては以下のような点が挙げられる。

- 5, 6 モーラなどの多モーラの語では複合語が多く、名詞の複合規則によって推定ができる。これに対して、3, 4 モーラ語のアクセント型は規則による推定が困難である。
- 3, 4 モーラ語は漢字 2 字で構成される場合が多い。
- 漢字 2 字の単語の場合、単語を構成する漢字がその単語のアクセント型に影響を与えると考えられる。

研究に用いたデータの数とそのアクセント型分布を表 1 に示す。

表 1: 研究対象データ

モーラ数	0型	1型	2型	3型	4型	計
3mora	4849	3735	275	135	-	8994
4mora	12889	957	531	403	30	14810

0型, 1型などのアクセント型は単語中のアクセント核(高低の境目となるモーラ)の位置を示し、3モーラ語ではそれぞれ0型(平板型), 1型(頭高型), 2型

(中高型), 3型(尾高型)に, 4モーラ語では0型(平板型), 1型(頭高型), 2型(中高型), 3型(中高型), 4型(尾高型)に対応する. 表2に3, 4モーラ語のアクセント型の例を示す.

表 2: 3, 4モーラの名詞のアクセント型の例

	0型	1型	2型	3型	4型
3モーラ					
4モーラ					

● アクセント線となるモーラ
○ モーラ

表 3: 統計データ (3モーラ) の例

	漢字	読み	総数 N_{all}	0型			1型			2型			3型		
				N_0	N_1	N_2	N_0	N_1	N_2	N_3	N_0	N_1	N_2	N_3	
語頭	家	カ	44	28	15	1	0								
	大	おお	4	2	2	0	0								
	大	タイ	42	4	37	0	1								
語尾	家	ガ	26	3	23	0	0								
	日	ジツ	10	2	8	0	0								
	夜	ヤ	25	2	22	1	0								

表 4: 統計データ (4モーラ語) の例

	漢字	読み	総数 N_{all}	0型		1型		2型		3型		4型	
				N_0	N_1	N_2	N_3	N_4	N_0	N_1	N_2	N_3	N_4
語頭	大	タイ	129	114	9	0	5	1					
	平	ヘイ	26	26	0	0	0	0					
	要	ヨウ	25	22	0	0	3	0					
語尾	願	カ	6	5	0	1	0	0					
	日	ジツ	41	37	3	1	0	0					
	板	ハン	14	14	0	0	0	0					

3 分析

3.1 構成漢字の種類と位置

先述したように漢字2字の単語の場合, 単語を構成する漢字がその単語のアクセントに影響を与えると考えられる. そこで単語単位のアクセント型の推定を行なう前に, まず各語を構成する漢字について, 漢字の種類と出現位置によって単語のアクセント型がどのような傾向になっているのかを定量的に分析するために, 研究対象データ中で語頭, 語尾の各漢字となりうる全ての漢字の種類について, 以下の4点に着目して統計データを各モーラ別に作成する.

- 漢字の表記
- 漢字の読み
音読み(片仮名で表す), 訓読み(平仮名で表す)
- 対象データ中で当該漢字を語頭(語尾)漢字として含む単語の総数
- 対象データ中で当該漢字を語頭(語尾)漢字として含む単語のアクセント型の分布(0型.. n型){ $n = 3(3モーラ), n = 4(4モーラ)$ }

ただしここで言う漢字の種類とは漢字の表記と読みを含めたものである. 表記が同じでも読みが異なれば, 異なる漢字として扱われる. 作成した統計データの一部を表3,4に示す.

3.2 アクセント法則

1章で述べたように, アクセントの法則とは単語についての単語の読みなどの音韻的な情報や単語の品詞, 活用, 素性, あるいは漢語, 和語, 外来語といった語の戸籍などの言語的な情報を基にしてその単語のアクセントにどのアクセント型になることが多いなどの傾向をもたせる法則である. これらの法則は以下のように表現することができる.

if (条件部) → アクセント型傾向

条件部にはアクセント法則を適用する条件が記述される. またアクセント型傾向はどの程度の割合でどのアクセント型になるといったことを示す. 文献[3], や[4]に記載されている法則の一部を例として以下に示す.

- 法則1
if (一般名詞, 漢語, 後部漢字1字2モーラ, 前部漢字1字1モーラ, 名詞的な意味を持つ)
→ 頭高型が多い
- 法則2
if (一般名詞, 漢語, 後部漢字1字2モーラ, 前部漢字1字1モーラ, 動詞的な意味を持つ)
→ 平板型が多い
- 法則3
if (一般名詞, 漢語, 後部漢字1字1モーラ, 前部漢字1字2モーラ, 名詞的な意味を持つ)
→ 頭高型が多い

• 法則 4

if (一般名詞, 漢語, 後部漢字 1 字 1 モーラ, 前部漢字 1 字 2 モーラ, 動詞的な意味を持つ)
→ 平板型が多い

これらの法則の多くでは, アクセント型傾向は「頭高型が多い」, 「多く平板型」などといったあいまいな表現がなされている。したがってこれらの法則が適用される単語がどのようなアクセント型の傾向になっているのか定量的に明らかにする必要がある。そこで研究対象データ中でこれらの法則が適用される語についてアクセント型の統計データを法則別に作成した。作成した統計の例を 5 に示す。

表 5: アクセント法則が適用される単語のアクセント型分布

法則	単語数	0 型	1 型	2 型	3 型
法則 1	2071	411	1634	17	9
法則 2	710	315	382	5	8
法則 3	2006	1465	454	87	0
法則 4	704	647	50	7	0

3.3 アクセント型傾向

前節でも述べたようにアクセント型傾向とはどの程度の割合でどのアクセント型になるといったものであり, アクセント辞書中では「頭高型が多い」, 「多数が平板型」といった表現がなされている。語頭, 語尾としてある漢字を含む単語あるいはアクセント法則が適用できる単語のアクセント型をこれらの傾向から推定する場合, 傾向を定量的に扱う必要がある。そこでまず, それら単語のアクセント型が n 型となる確率を n 型アクセント率とし以下のように求める。

3.1, 3.2 節で作成した統計データ統計データはそれぞれ漢字別, 法則別に, 単語の総数 (N_{all}) とそれら単語のアクセント型の分布 (N_0, N_1, \dots) で構成されている。したがって n 型アクセント率を P_n と表わすと, P_n は式 1 により定義できる。

$$P_n = \frac{N_n}{N_{all}} \quad (1)$$

N_{all} 単語の総数
 N_n 属する単語の内アクセント型が n 型の単語の総数
 P_n n 型アクセント率

アクセント型傾向は単語が 0 型, 1 型, n 型となるそれぞれの度合を示すので, これは n 型アクセント率の

$n = 0, 1, \dots, N$ (但し $N = 3$ (3 モーラ語), $N = 4$ (4 モーラ語)) の集合 $\{P_0, P_1, \dots, P_N\}$ で表わせる。

4 アクセント型推定手法

4.1 推定手法の流れ

前章では単語のアクセントに語頭, 語尾漢字の種類や位置, あるいはアクセント法則がどのようなアクセント型の傾向を単語に与えるのか分析した結果

- 語頭漢字の種類によるアクセント型傾向
- 語尾漢字の種類によるアクセント型傾向
- アクセント法則によるアクセント型傾向

を得た。ここで図 1 に示すようにこれらのアクセント型傾向と, アクセント法則によるアクセント型傾向を結合し新たなアクセント型傾向を生成する。

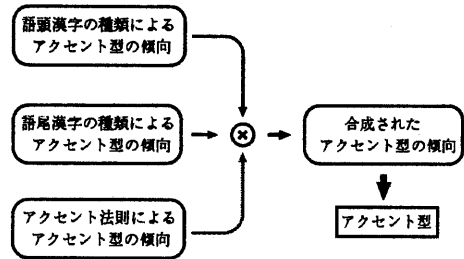


図 1: 推定手法

生成されたアクセント型傾向のうち傾向が最も強い型が推定される型となる。

4.2 アクセント型傾向の結合

図 1 に示すされたようにアクセント型の傾向は, それぞれ独立な証拠から得られる。これらを旨くまとめるために Dempster の結合規則を用いる。

Dempster の結合規則は独立な証拠から推論された基本確率を結合する法則である [10]。例えば, m_1, m_2 を独立な証拠に基づいて得られた基本確率とし $A_{1i}, A_{2,j} (i, j = 0, 1, 2, \dots)$ をそれぞれの確率における焦点の要素とする。この際に $A_{1i}, A_{2,j}$ の積集合 A_k の基本確率は式 2 によって定義される。

$$m(A_k) = \frac{\sum_{A_{1i} \cap A_{2j} = A_k} m_1(A_{1i})m_2(A_{2j})}{1 - \sum_{A_{1i} \cap A_{2j} = \phi} m_1(A_{1i})m_2(A_{2j})} \quad (2)$$

(但し $A_k \neq \phi$)

三つ以上の基本確率を結合する場合、これまでに結合した二つの基本確率と他の基本確率を式2によって結合すればよい。

4.3 推定の例

例として3モーラの一般名詞「果実」(カジツ)のアクセント型推定を試みる。推定手順としては以下の通りである。

まず3.2節で作成した統計データから「果実」の語頭漢字“果”(カ)と語尾漢字“実”(ジツ)をそれぞれ検索する。各漢字の統計データを表6に示す。また果実は3.2で示した法則1の条件部を満す語なので統計データから法則1のデータも検索しておく。(表5参照)

表6: 「果実」の構成漢字の統計データ

位置	漢字	読み	総数	0型	1型	2型	3型
				N_{all}	N_0	N_1	N_2
語頭	果	カ	8	7	1	0	0
語尾	実	ジツ	9	3	6	0	0

各漢字の統計データから各漢字によるアクセント型傾向とアクセント法則(法則1)によるアクセント型傾向を3.3節に示した手法によって求める。求めた傾向を表7に示す。

表7: 「果実」のアクセント型傾向

		0型	1型	2型	3型
		P_0	P_1	P_2	P_3
(1)	語頭漢字(果)	0.875	0.125	0.000	0.000
(2)	語尾漢字(実)	0.333	0.666	0.000	0.000
(3)	法則1	0.198	0.789	0.008	0.004

次にこれらを Dempster の結合法則(式2)により結合する。結合した結果を表8に示す。

表8: アクセント型傾向の結合結果

	0型	1型	2型	3型
	P_0	P_1	P_2	P_3
(1)+(2)	0.778	0.222	0.000	0.000
(1)+(2)+(3)	0.466	0.531	0.000	0.000

この場合、語頭、語尾漢字のアクセント型の結合結果は表8中の(1)+(2)に示すようになっている。この結果からアクセント型を推定すると傾向が最大となっている0型が推定結果となる。文献[3]に示されている「果実」のアクセント型は1型であるのでこれは誤りである。一方アクセント法則によるアクセント型傾向を取り入れて結合した結果は表8中(1)+(2)+(3)に示すようになっている。この結果より推定型は1型となり文献[3]に示されている「果実」のアクセント型と一致する。

5 推定結果

5.1 アクセント法則を取り入れない場合

3, 4モーラ漢字2字の一般名詞に対し、構成漢字の種類と位置によるアクセント型傾向を考慮することによってアクセント型の推定を行なった結果を表9に示す。

表9: アクセント型推定結果1

モーラ数	データ数(語)	正解数(語)	正解率(%)
3mora	8994	8064	89.7
4mora	14810	13652	92.2

5.2 アクセント法則を取り入れる場合

5.1節では、構成漢字の種類と位置によるアクセント型傾向を考慮することによって推定を行なった結果を示した。本研究ではこれらの傾向に加えてアクセント法則によるアクセント型傾向を取り入れて推定を行なった結果を表10に示す。ただしここで取り入れたアクセント法則は3.2節で例として示したものである。これらは3モーラ語に対する法則ばかりなので推定の対象は3モーラ語のみとした。

表10: アクセント型推定結果2

モーラ数	データ数(語)	正解数(語)	正解率(%)
3mora	8994	8119	90.3

6 考察

5.1, 5.2節では3モーラの一般名詞のアクセント型の推定にアクセント法則を取り入れた場合と取り入れなかった場合の正解率を示した。その結果正解率はアクセント法則を取り入れた場合のほうが0.6%高いことがわかった。それぞれの場合の推定結果のアクセント型分布を表11, 12にそれぞれ示す。

これらの表より推定されたアクセント型の分布をそれぞれの型で比較してみると、アクセント法則を取り入れた場合のほうが1型の不正解数が少ないが、逆に0型の不正解数は増加していることがわかる。2型, 3型に関しては不正解数にほとんど変化はないが、推定結果2または3型となる単語数が減少している。

これは3.2節に示した法則が0または1型の傾向が強いものばかりであったためと考えられる。したがって2型, 3型になる傾向が強い法則を取り入れることにより改善されると考えられる。

表 11: 3モーラ語のアクセント型分布 (アクセント法則なし)

	0型	1型	2型	3型
対象データ	4849	3735	275	135
推定結果	5039	3702	170	83
正解	4496	3322	165	81
不正解	543	380	5	2

表 12: 3モーラ語のアクセント型分布 (アクセント法則あり)

	0型	1型	2型	3型
対象データ	4849	3735	275	135
推定結果	5167	3614	139	74
正解	4585	3318	134	72
不正解	582	296	5	2

7 まとめと今後の課題

本報告では語頭, 語尾の漢字の種類によるアクセント型傾向によってアクセント型を推定する手法を基本とした上で, アクセント法則によるアクセント型傾向を取り入れることのできるアクセント推定手

法を提案し, 推定を行なった。その結果アクセント法則を取り入れることによって, 推定の結果を向上できることがわかった。今回推定に取り入れた法則は言語的な情報を基にした法則で, 実際の推定に取り入れることのできたアクセント法則はまだわずかである。これは法則を適用する条件が言語的な情報であるため判断が難しいことが原因である。例えば,

if (一般名詞, 漢語, 後部漢字1字2モーラ, 前部漢字1字2モーラ, 古くからの語) → 頭高型が多い

といった法則では条件部は「古くからの語である」となるが, 語が古いか古くないかの判断は非常に難しい。したがって今後は比較的条件的判断が容易である音韻的な情報を基にしたアクセント法則を積極的に取り入れて推定を行なって行きたい。

参考文献

- [1] 広川, 佐藤: “人名(姓)におけるアクセント型推定法”, 音講論, pp.425-426(1981.5)
- [2] 北川, 樋口: “日本人の姓のアクセント型推定”, 音講論, pp.227-228(1996.9)
- [3] 安藤達夫: “日本語 発音アクセント辞典”, 日本放送出版協会
- [4] 金田一 春彦: “明解日本語アクセント辞典”, 三星堂
- [5] 春彦: “広辞苑”, 三星堂
- [6] 杉藤 美代子: “「大阪・東京アクセント音声辞典」CD-ROM 解説編”, 丸善株式会社
- [7] 杉藤 美代子: “日本語と日本語教育 3 日本語の音声・音韻(下)”, 明治書院
- [8] 玉村 文郎: “日本語と日本語教育 6 日本語の語彙・意味(上)”, 明治書院
- [9] 杉藤 美代子: “日本語の音”, 和泉書院
- [10] 石塚 満: “Dempster&Shafer の確率理論”, 電子通信学会誌, pp.900-903(1983.9)