社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

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利用した統計的アクセント結合処理の検討

黒岩　龍†　峯松　信明††　伝　康晴†††　広瀬　啓吉†

† 東京大学大学院情報理工学系研究科
〒 113–8656 東京都文京区本郷 7–3–1
†† 東京大学大学院新領域創成科学研究科
〒 277–8561 千葉県柏市柏の葉 5–1–5
††† 千葉大学文学部
〒 263–8522 千葉県千葉市弥生町 1–33
E-mail: †{kuroiwa,hirose}@gavo.t.u-tokyo.ac.jp, ††mine@k.u-tokyo.ac.jp, †††den@cogsci.l.chiba-u.ac.jp

あらまし　日本語テキスト音声合成では，単語連鎖に伴うアクセント変化（アクセント結合）を適切に処理する必要がある。従来筆者らは規則を用いてアクセント結合処理を実装してきたが，誤った出力をする例が少なからず見られる。そこで，本研究では，まず 1 人のみのラベラによる大規模なアクセントデータベースを作成した上で，そのデータベースを使用して条件付確率場（CRF）による統計的学習・推定を行ない，アクセント結合処理を統計的手法で実装した。さらに，統計的手法に従来の規則から得られる知見を導入することで，精度を向上させた。その結果，従来手法や単純な統計的処理と比して，誤った出力の率を大きく減らすことに成功した。
キーワード　アクセント結合, テキスト音声合成, データベース, 統計的学習, 条件付確率場

Accent Labeling of a Large-scale Database by a Single Labeler and its Use in Statistical Learning of Accent Sandhi

Ryo KUROIWA†, Nobuaki MINEMATSU††, Yasuharu DEN†††, and Keikichi HIROSE†

† Graduate School of Information Science and Technology, The University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
†† Graduate School of Frontier Sciences, The University of Tokyo
5–1–5 Kashiwanoha, Kashiwa-shi, Chiba, 277–8561 Japan
††† Faculty of Letters, Chiba University
1–33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263–8522 Japan
E-mail: †{kuroiwa,hirose}@gavo.t.u-tokyo.ac.jp, ††mine@k.u-tokyo.ac.jp, †††den@cogsci.l.chiba-u.ac.jp

Abstract　For developing a Japanese TTS system, it is important to realize a module of word accent sandhi adequately. Our previous works developed the module by using rules but it is a fact that the rule-based module produced an unignorable number of errors. In this report, accent labeling of a large-scale corpus was done by a single labeler and a corpus-based module for word accent sandhi was developed, where conditional random fields (CRF) were used. The new module was improved further by integrating the rule-based knowledge on accent sandhi. Experimental results showed that the method proposed by integrating corpus-based and rule-based modules outperformed the conventional rule-based method and the quickly-developed corpus-based method.
Key words　accent sandhi, text-to-speech, database, statistical learning, conditional random fields

1. はじめに

日本語テキスト音声合成システムを構築する場合，一般的には，次のような言語処理系／波形生成系が必要となる。1) 形態素解析を行ない，形態素境界を特定し，各形態素の読みやアクセント型（各形態素を単独で読み上げた際のモーラ列・音素列とアクセント型）の情報を得る，2) 無声化，連濁などの音韻処理，アクセント結合・イントネーション・継続長・パワーに関する韻律処理を行なう，3) 上記の情報を基に，波形生成を行なう。

従来筆者らは，アクセント句境界が与えられた場合に，句内のどのモーラをアクセント核として波形生成するのか（アクセント結合問題）を，各形態素（自立語／付属語）のアクセント属性を定義し，規則によってアクセント変化を記述することでシステム構築を行なって来た [1], [2]。しかし，全ての事象を規則で網羅することには限界があり（例えば [1] では副次アクセントや，付属語連鎖への対処が行なわれていない），アクセントラベルが施された大規模なコーパスを用いた機械学習・統計学習で解決を図ることも行なわれるようになった [3]。

コーパスベースの方法論を検討する場合，当然，大規模コーパスは必須のものとなるが，現時点で，高品質のアクセントラベリングが施され，研究目的で自由に利用可能なコーパスは存在していない。これらの現状を鑑み本研究では，1) アクセント結合処理モジュールを構築可能な，大規模かつ高品質なアクセントラベリングが施されたコーパスの構築と，2) それに基づく（かつ，従来の規則ベースのアクセント結合処理を踏まえた）統計的なアクセント結合処理モジュールの構築を試みる。

2. 特定ラベラによる大規模アクセントラベリングコーパス

2.1 ラベリング対象とする言語事象

日本語東京方言で文を発声する際，文は幾つかのまとまりに分かれ，各々の内部では音の高さ（ピッチ）が連続的に変化する。まとまりが始まる箇所ではピッチの上昇が見られ，その後，まとまりの内部では上昇は無く，ゆるやかに下降してゆく。まとまりの内部には，語彙に依存した比較的急激なピッチの下降箇所が概ね高々1 つ存在する。このようなまとまりをアクセント句と呼び，急激な下降の箇所をアクセント核と定義するのが一般的である。

しかし，実際の発声におけるピッチ変化を観察すると，明確にアクセント句の分離が困難な場合も多く，複数の句が影響を及ぼし合い，融合する現象も見られる。これについて，日本語話し言葉コーパス [4] では，後続アクセント句の句頭上昇が見られない程度まで融合が起きていれば 1 つのアクセント句として扱うものとするが，同程度の融合が見られても双方が核を持つアクセント句である場合は,「アクセント句には 1 つの核しか存在し得ない」との大前提に従い，複数のアクセント句として扱っている。

アクセント核についても，発声者や聴取者にとって知覚されているにも拘らず，実際のピッチ変化に明確に現れていない場合がある [4]。更には，発声者等の知覚も必ずしも明確ではない。

このようにアクセント句・核は明確に（物理的に）定義できるものでないが，ラベリング作業を行なう場合，何らかの定義が必要となる。本研究では下記の言葉でこれらを定義し，ラベリング対象とした。

アクセント句境界　ピッチの句頭上昇が見られる箇所をアクセント句境界とする。視覚提示される話速（約 7 モーラ/秒）に合わせて自然に読んだ場合を想定する。休止が入った場合も，通常は境界が生じる。

アクセント核　ピッチが急激に下降する箇所の直前のモーラ。句内の出現回数は制限しない。意識的に当該箇所で急激にピッチを下げ，それ以外では同じピッチで平坦に（イントネーションを除去して）読んだ場合に，違和感が生じない位置。

上記のラベリング対象は，文読み上げ時の事象である。これとは別に，個々の自立語を単独で発声する場合のアクセント型もその対象とした。後述するように本ラベリングは特定のラベラによる作業となる。単独発声時のアクセント型は，アクセント辞典等に掲載されているが，アクセント感覚の個人差を含まない高品質のコーパス構築を念頭に置き，単独発声時に対するラベリングも作業項目とした。

最終的に，本研究におけるラベリング対象は，1) 文発声時のアクセント句境界とアクセント核位置，及び，2) 文中の全自立語に対する単独発声時のアクセント核位置，である。なお，付属語については，単独発声を想定すること自体が困難であり，また，不合理とも考えられるため，これらに対する単独発声時のアクセントラベリングは行なわない。

2.2 ラベラ・ラベル検査者の選定

アクセント感覚には個人差があるため，本研究では特定ラベラに全ラベリングを依頼することとした。作業量が膨大となるため，誤ラベリングも免れ得ない。そこで，付与されたラベルを検査する（誤りが含まれる可能性のある文を選定する）検査者をラベラとは別に用意した。東京生まれ・育ちであり，合唱部に所属する比較的音感の鋭い大学生 6 名に対して，日本語アクセントに対する教育を施した上で，試験により選抜し，最終的にラベラ 1 名，検査者 1 名（今後増員する可能性あり）を選定した。

2.3 使用した文セット

新聞記事読み上げ音声コーパス（JNAS）で使用されている文（毎日新聞記事から抽出した 16,178 文および ATR 音素バランス文 503 文）を用いた。既に音声コーパス用に使用されている文であり，全文に凡そ正しい読みが与えられていること，砕けた文が少なく扱いやすいことなどが選定理由である。但し，不自然な読みや誤った読みも散見されたため，事前に全文の読みを確認し，適切でないと思われる箇所に対しては，適宜，修正を施した。

2.4 形態素解析

文中の自立語に対するラベリング作業を行なうこと，及び，ラベリング結果を利用する際の利便性のため，全文に対して形態素解析を行なった。形態素解析の品詞体系は UniDic に準拠

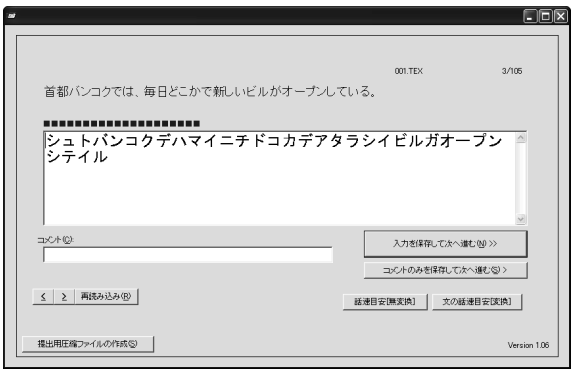

図 1　ラベリング作業用インタフェース

表 1　アクセント句を構成する形態素の数

形態素数	出現数	出現率
1	5079	17.4%
2	9829	33.6%
3	7902	27.0%
4	3972	13.6%
5	1586	5.4%
6	554	1.9%
7 以上	303	1.0%

した。形態素解析作業は，辞書として UniDic を用いて自動的に行なった後，用意した読みと異なった読みが付与されたものを中心に手修正した。従って，全て正確に形態素解析が行なわれている訳ではないが，読みについてはラベリング作業に用いたものと完全に同一になっている。

2.5　実際のラベリング作業

ラベリング作業では，音声の録音はせず，テキストに直接ラベルを付与させた。これは，音声を経ると極めて作業時間が長くなるためであり，また，ラベラの言語的内省としてのアクセント情報を引き出すことを重視するためである。作業用に，図 1 に示したインタフェースを用意し，単純な操作でラベリング作業ができるようにした。

単独発声時の自立語のラベリングでは，完全な単独発声では曖昧性が残るため，名詞には助詞「が」を後続させた状態でラベリングをさせ，形容詞には名詞「こと」が後続することを想定してラベリングをさせる（「こと」自体はラベリングの対象としない）などの工夫を行なった。この作業により，

- **シュ ’ト／バ ’ンコクデ ’ハ／マ ’イニチ／ド ’コカデ／アタラシ ’イ／ビ ’ルガ／オ ’ープン／シテイル**

のような文発声ラベリングと，

- **シュ ’トガ**
- **アタラシ ’イ**
- **スル**

のような形態素単独発声ラベリングが得られる。なお，［／］がアクセント句境界位置を，［ ’］がアクセント核位置を表している。ラベル検査者から誤りの可能性を指摘された文については，ラベラに確認させ，必要に応じて，訂正させた。

2.6　進捗状況と分析

2007 年 1 月現在，4,166 文（JNAS の先頭 40 ファイル）について，文発声・形態素単独発声の双方のラベリングが完了している。今後，用意した全ての文に対してのラベリングを目指して進めていく。作業が完了した全文に対して，アクセント句を構成する形態素数，及び，アクセント句を構成する品詞列の上位 10 種類を表 1，表 2 に示す。4 形態素以下のアクセント句が 90%以上を占めていることや，品詞列は 11 位以下の低い出現率のものの合計がおよそ半数に達することなどが読み取れる。以降の節では，構築したコーパスを用いたアクセント結合処理モジュールについて検討する。

表 2　アクセント句を構成する品詞列（上位 10 種）

品詞列	出現数	出現率
[名][助]	5273	18.0%
[名]	2639	9.0%
[名][名][助]	2180	7.5%
[名][接尾][助]	1409	4.8%
[動][助動]	792	2.7%
[動]	788	2.7%
[名][名]	758	2.6%
[名][接尾]	739	2.5%
[動][助]	571	2.0%
[名][助][助]	541	1.9%
上記以外	13535	46.3%

3.　CRF を用いたアクセント結合

3.1　条件付確率場

観測データ $\boldsymbol{x}$ に対する出力ラベル $\boldsymbol{y}$ を学習するに際し，条件付確率場（CRF）[5] は $(\boldsymbol{x}, \boldsymbol{y})$ 内での連続する変数の組（y_{t-1} と y_t, y_t と x_t など）の関係についての独立した特徴（素性）f を列挙し，各素性 f の重要度を θ_f，$(\boldsymbol{x}, \boldsymbol{y})$ 内で素性 f が満たされている箇所の数を $\phi_f(\boldsymbol{x}, \boldsymbol{y})$ とおいた上で，入力 $\boldsymbol{x}$ に出力 $\boldsymbol{y}$ を割当てることの確信度として，$\sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y})$ を考え，これを，

$$\Pr(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y})}{\sum_{\boldsymbol{y} \in Y} (\exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y}))}$$

として確率分布とする。正解データを与えることによる学習は，この確率をできるだけ大きくするような重要度 θ_f を探る作業となる。本研究では，CRF++ [6] を用いてアクセント結合処理を実装した。

3.2　学習・推定の内容

アクセント句境界は事前に与えられているものとし，各句中のアクセント核位置について学習・推定した。これらの情報は，前節で構築したコーパスより得られる。アクセント句境界位置情報を使用してアクセント句に区切り，句に含まれる各形態素の文中アクセント型を学習・推定の対象とした。例えば「音声合成」であれば「オンセー」（無核）と「ゴーセー」（1 モーラ目に核）という文中アクセント型を，単独発声時の「音声（オンセー）」と「合成（ゴーセー）」から推定する枠組みである。なお，形態素境界に核が生じているもの（「流れ・を（ナガレ・オ）」「出来・ない（デキ・ナイ）」など）では，当該アクセン

	A	B	C	D	E	F	G	H	I	J	K
1	まだ	マダ	未だ	マダ	副詞	*	*	マ’ダ／	2	1	1
2											
3	正式	セイシキ	正式	セイシキ	名詞-普通名詞-形状詞可能	*	*	セイシキ	4	-1	-1
4	に	ニ	だ	ダ	助動詞	助動詞-ダ	連用形-ニ	ニ／	1	*	-1
5											
6	決まっ	キマッ	決まる	キマル	動詞-一般	五段-ラ行-一般	連用形-促音便	キマッ	3	-1	-1
7	た	タ	た	タ	助動詞	助動詞-タ	基本形-一般	タ／	1	*	-1
8											
9	わけ	ワケ	訳	ワケ	名詞-普通名詞-一般	*	*	ワ’ケ	2	1	1
10	で	デ	で	デ	助詞-格助詞	*	*	デ’	1	*	1
11	は	ハ	は	ハ	助詞-係助詞	*	*	ハ／	1	*	-1
12											
13	ない	ナイ	無い	ナイ	形容詞-非自立可能	形容詞-ア段-無イ+ない	基本形-一般	ナ’イ	2	1	1
14	の	ノ	の	ノ	助詞-準体助詞	*	*	ノ’	1	*	1
15	で	デ	だ	ダ	助動詞	助動詞-ダ	連用形-一般	デ	1	*	-1
16											
17	カネ	カネ	金	カネ	名詞-普通名詞-一般	*	*	カネ	2	-1	-1
18	の	ノ	の	ノ	助詞-格助詞	*	*	ノ	1	*	-1
19	力	チカラ	力	チカラ	名詞-普通名詞-一般	*	*	チカラ’	3	3	3
20	も	モ	も	モ	助詞-係助詞	*	*	モ／	1	*	-1
21											
22	十分	ジュウブン	十分	ジュウブン	副詞	*	*	ジュウブ’ン／	4	3	3
23											
24	知っ	シッ	知る	シル	動詞-一般	五段-ラ行-一般	連用形-促音便	シッ	2	-1	-1
25	て	テ	て	テ	助詞-接続助詞	*	*	テ	1	*	-1
26	い	イ	居る	イル	動詞-非自立可能	上一段-ア行	連用形-一般	イ	1	-1	-1
27	た	タ	た	タ	助動詞	助動詞-タ	基本形-一般	タ	1	*	-1
28											
29	首都	シュト	首都	シュト	名詞-普通名詞-一般	*	*	シュ’ト／	2	1	1
30											
31	バンコク	バンコク	バンコク	バンコク	名詞-固有名詞-地名-一般	*	*	バ’ンコク	4	1	1
32	で	デ	で	デ	助詞-格助詞	*	*	デ’	1	*	1
33	は	ハ	は	ハ	助詞-係助詞	*	*	ハ／	1	*	-1

図 2　形態素解析結果と対応付けたアクセントラベル

ト核は前側の形態素に属するものとした。このようにして形態素解析結果，単独発声/文中アクセント型をまとめた結果が図 2 であり，列 H が文中アクセントラベル済みテキストを形態素に対応付けたもの，列 K が形態素の文中アクセント型，列 J が形態素の単独発声アクセント型である。なお，“−1” はアクセント核が存在しないことを表し，“*” は単独アクセント型ラベリングの対象となっていない形態素であることを表す。その他の数値は核のあるモーラの位置を表している。

但し，形態素内にアクセント句境界が存在している文や複数のアクセント核を持つ形態素を含む文は文ごと除去した。最終的に，学習用 3,581 文（25,692 アクセント句），推定用 527 文（3,533 アクセント句）に分けて使用した。JNAS の 40 ファイルを，35 ファイルと 5 ファイルに分割した。

3.3　単独アクセント型を与えない学習

形態素単独発声アクセント型が与えられていない状況（[3] と同様の条件）を想定し，観測素性として，前後 2 形態素を含めた 5 形態素について，[基本形/基本形読み/書字形/品詞/活用型]（以下，“基本形等” と記す。），[品詞]，[品詞 (大分類のみ)]，[活用型 (大分類のみ)]，[活用形 (大分類のみ)]，[モーラ数] のそれぞれと当該形態素の文中アクセント型との関係を与えた。また，遷移素性として，連接する形態素の文中アクセント型同士の関係を与えた。CRF++で学習・推定した結果が表 3 上段である。82.1% のアクセント句について正しい推定が得られている。{ 名詞, 動詞, 形容詞, 形状詞 }+{ 助詞, 助動詞 } の 2 語で構成された単純なアクセント句を対象とした場合は正解率が高く（85.9%），逆に名詞の連続（複合名詞）を含む句では率が低くなる（77.0%）。なお，複数の核を持つアクセント句については，最初の核（主アクセント）が一致していれば正解と見なした。全ての核が一致した場合のみを正解とすると，最大で 3%程度正解率が低下する。

3.4　単独アクセント型を与える学習

前節での素性に加え，観測素性に [単独発声アクセント型] を加えて同様の学習を行なった。推定結果は，表 3 の 2 段目に示している。単独型未使用時と使用時では多くの違いが見られ，特に単純な句においては 85.9% から 94.3% になるなど顕著な違いが見られる。しかし，複合名詞を含む句では正解率に向上が見られない。単純な句では単独型がそのまま文中型となることが多いのに対し，複合名詞においては型変形が頻出することによると考えている。

3.5　簡易変化ラベルの学習

以上の型推定は，文中アクセント型を直接学習・推定の対象としていたため，単独発声型と文中型がいずれも “1” である場合と “2” である場合は別々の事象として扱われる。そこで，単独型から文中型への「型変化」の様子を学習・推定の対象とすることで，類似する現象を共通のものと捉えられるようにした。具体的には次のようなラベル（“簡易変化ラベル” とする。）を使用して学習・推定を行なった。

単独型が有核の場合，文中型が有核であれば，単独型からの変化量を表すラベル（“[0]”; “[+1]”, “[+2]”, ...; “[−1]”, “[−2]”, ...）を学習対象とし，文中アクセント型が無核であれば，無核を示すラベル（“non”）を学習対象とした。一方，単独型が無核あるいは，値を持たないものに対しては，文中型を直接の学習対象とした。推定結果より，機械的に文中型に相当する数値に復元する。表 3 の 3 段目に結果を示す。

単独型から文中型への型変化を推定の対象とした場合では，文中型を直接推定する場合と比較し，全体的に正解率の向上が見られる。相対的な変化を学習の対象とすることで，効率の良い学習ができたものと言える。

3.6　隣接形態素組み合わせ素性を用いた学習・推定

ここまでに用いた観測素性は，当該形態素または周辺形態素

表 3　CRF による文中アクセント型推定の結果

	形態素単位		アクセント句単位					
			すべての句		単純な句		複合名詞を含む句	
直接学習（単独型なし）	9080 /9908	91.6%	2833 /3533	82.1%	703 /822	85.9%	530 /688	77.0%
直接学習（単独型あり）	9272 /9908	93.6%	3081 /3533	87.2%	775 /822	94.3%	523 /688	76.0%
簡易変化ラベルの学習	9319 /9908	94.1%	3137 /3533	88.8%	791 /822	96.2%	553 /688	80.4%
簡易変化ラベルの学習（組合せ素性使用）	9424 /9908	95.1%	3214 /3533	91.0%	790 /822	96.1%	578 /688	84.0%
規則適合変化ラベルの学習（組合せ素性使用）	9458 /9908	95.5%	3238 /3533	91.7%	792 /822	96.4%	589 /688	85.6%

表 4　CRF による文中アクセント型推定の結果（許容度 3 以上を正答とする場合）

	アクセント句単位					
	すべての句		単純な句		複合名詞を含む句	
規則適合変化ラベルの学習（組合せ素性使用）	3307 /3533	93.6%	808 /822	98.3%	605 /688	87.9%

についての品詞等のそれぞれと出力ラベルの関係のみであった。結果として, [1] の複合単語アクセント結合規則のように，名詞が連続した場合にのみ起こる現象，例えば，当該形態素が名詞で 1 つ後の形態素も名詞である場合に文中アクセント型が無核になることを学習する場合，2 つの観測素性

- 当該形態素の品詞が名詞である場合には無核になる
- 1 つ後の形態素の品詞が名詞である場合には無核になる

についての重要度 θ_f を上げる処理がされる。しかし，これは名詞の連続以外に対しても影響を与えてしまい，不都合である。そこで，特定の組み合わせに特化した学習のために，これまでの観測素性に加え，[当該形態素の品詞/**前**の形態素の品詞], [当該形態素の品詞/**後**の形態素の品詞], [当該形態素の品詞, **前**の形態素の基本形等], [当該形態素の品詞, **後**の形態素の基本形等], [当該形態素の基本形等, **前**の形態素の品詞], [当該形態素の基本形等, **後**の形態素の品詞] のそれぞれと当該形態素の文中アクセント型との関係を観測素性として使用し，相対変化ラベルの学習・推定をした。品詞と品詞の組み合わせ以外にも，当該形態素の品詞と前後形態素の基本形等，当該形態素の基本形等と前後形態素の品詞について組み合わせることで，例外的なアクセントについても学習がされやすいように配慮した。その結果が表 3 の 4 段目である。名詞連続を含むアクセント句について特に顕著な改善が見られ，組み合わせの学習が効果を発揮していると言える。

3.7　アクセント結合規則に適合させた学習・推定

以上のように，アクセントの変化を学習の対象とするなど, [1] の規則で可能な処理と類似した結果が得られる学習を行なうことで，高い正答率が得られた。よって，規則で表現されている処理をさらに忠実に取り入れた形で学習することにより，規則的現象をより的確に捉えることができるものと考えられる。そこで, [1] の規則を基に検討し，以下のような相対変化ラベルを導入し，これを学習・推定の対象とすることにした。

- 単独発声アクセント型が有核のもの（“−1” 以外の値を持つもの）に対しては，1) 文中アクセント型が無核（“−1”）の場合，その旨を示すラベル（“non”），2) 文中アクセント型が単独発声型と同じである場合，その旨を示すラベル（“same”），3) 文中アクセント型がモーラ数と同じである（末尾に核がある）場合，その旨を示すラベル（“morae”），4) 文中アクセント型が単独発声型より 1 小さい場合，その旨を示すラベル（“same−1”），5) 文中アクセント型が 1 型の場合，その旨を示すラベル（“one”），6) 文中アクセント型がモーラ数より 1 小さい（末尾の 1 つ前に核がある）場合，その旨を示すラベル（“morae−1”），7) その他の場合は，単独発声型からの変化量を表すラベル（“[0]”; “[+1]”, “[+2]”, . . .; “[−1]”, “[−2]”, . . .）
- 単独発声アクセント型が無核のもの（“−1”）に対しては，1) 文中アクセント型が単独型と同様に無核（“−1”）の場合，その旨を示すラベル（“samenon”），2) 文中アクセント型がモーラ数と同じである（末尾に核がある）場合，その旨を示すラベル（“morae”），3) 文中アクセント型が 1 型の場合，その旨を示すラベル（“one”），4) 文中アクセント型がモーラ数より 1 小さい（末尾の 1 つ前に核がある）場合，その旨を示すラベル（“morae−1”），5) その他の場合は，文中アクセント型（“1”, “2”, . . .; “−1”）
- 単独発声アクセント型を持たないものに対しては，文中アクセント型（“1”, “2”, . . .; “−1”）

これらは，アクセント変化の結果として発生しやすい現象を示すラベルから順に記しているため，複数に該当しうるアクセント句については，最も上にあるものを相対変化ラベルとして使用する。

また，学習・推定のデータに，1) 単独発声アクセント型が無核である場合，その旨を示すラベル（“non”），2) 単独発声アクセント型がモーラ数と同じである（末尾に核がある）場合，その旨を示すラベル（“morae”），3) 単独発声モーラ数より 1 小さい（末尾の 1 つ前に核がある）場合，基本形読みの末尾 2 モーラ（“オイ” など），4) 単独発声型が上記に当てはまらない場合，その旨を示すラベル（“else”）とするラベル（“単独型種類ラベル” とする）を用意し，単独発声アクセント型の種類による結果の分岐を行ないやすくすることを目指した。

また，学習・推定の際に使用する観測素性に，当該形態素についての [単独型種類ラベル], [出現形読みの 先頭のモーラ], [出現形読みの 第 2 モーラ], [出現形読みの 単独発声核位置の前のモーラ], [出現形読みの 単独発声核位置のモーラ], [出現形読みの 単独発声核位置の後のモーラ], [出現形読みの 末尾の前のモーラ], [出現形読みの 末尾のモーラ] のそれぞれと当該形態素の文中アクセント型との関係を加えた。出現形の読みの各モー

図 3　許容度判定プログラム

ラを用いたのは, [1] の規則における音節内移動則（アクセント核が音節の先頭のモーラに移動する現象）などの影響を学習させるためである。

上記のように，出力ラベルや観測素性に工夫をして学習・推定をした結果が，表 3 の 5 段目である。すべての場合に渡って，4 段目のものからの向上が見られる。

3.8　不正解に分類された形態素の許容度に関する調査

統計的なアクセント学習に各種の工夫を重ねることにより，アクセント推定は，91.7% のアクセント句に対して正しい結果が得られており（主アクセント核のみについて考える場合），単純なアクセント句に限れば正答率は 96.4% に及ぶ。しかし，このことは，単純なアクセント句であっても 3.6% の句ではアクセント推定結果と正解ラベルが一致しないことを意味する。それらの不一致が本質的な誤りであるか，アクセントの揺れの範囲内として許容できるものなのかによって，結果の意味するところは異なる。そこで，アクセント推定結果と正解ラベルが一致しなかったアクセント句について，アクセントデータベースを作成する際にラベリングを行なったラベラに，許容度を判定させた。

許容度の判定には図 3 のプログラムを使用し，3.7 節の実験において正答とならなかった（主アクセント核のみについて考える場合にだけ正答となったものを含む）アクセント句 374 句に対して，1（とても不自然）から 4（とても自然）の 4 段階で許容度を回答させた。この結果，3 以上の許容度を得られたものは正答と見做すこととすると，表 3 の下段のようになる。単純な結合では，98.3% ものアクセント句について正しい結果または許容できる結果が得られていることとなった。音声として出力する場合にはさらに許容できる幅が広がると考えられるため，完全に近い精度であると言うことができるだろう。ただし，許容できないと判定されたアクセント句がわずかながら存在したのは事実であり，何らかの方法で改善することが望まれる。

3.9　考　　察

統計的な学習にアクセント結合規則から得られる知識を導入し，正答率を高めることに成功した。

隣接する形態素の品詞を組にして観測素性に加えたことによる改善からは, [1] のアクセント結合規則で結合する品詞によって規則が使い分けられているように，品詞がどのように連続するかが文中でのアクセントを考える上で重要であることが分かる。その影響は、当該形態素や前後の形態素の品詞を個別に考えるのでは不十分となるほどに大きいものである。

アクセント結合規則に細かく適合させた学習において正答率の向上が見られたことからは，規則が完全であるとは言えないものの，アクセント結合処理の重要な部分を押さえていることが分かる。今回の実験では，アクセント結合規則への細かな適合をする前とした後のみの比較をしたため，どの部分が特に大きな影響を与えたのかについては不明であり，この点は今後の課題といえる。

また，各種手法による改善をしても，名詞の連続を含むアクセント句の約 12% では，推定結果に許容できない誤りがあり，改善が必要である。複合名詞では，文脈や係り受け関係によって結合の仕方に違いが出ることがあり，揺れも多い。例えば，「政権交代」が「セーケンコータイ」「セーケンコータイ」のいずれにもなり得るが,「配達証明」（ハイタツショーメー）,「合併失敗」（ガッペーシッパイ）のように一方しか許容されないものも多い。ここに掲げた複合名詞はいずれも単独発声では無核となる名詞で構成されているにもかかわらず結果が異なるため，これらを的確に判別するのは難しい。また,「ソ連邦解体後」「日米交換船」のようなアクセント句では係り受けの影響を受ける可能性が高い。このような問題は規則に基づくアクセント結合処理でも解決できていない問題であり，新たな手法が必要になると考えられる。

4.　ま　と　め

コーパスベース及び規則ベースのハイブリッド型アクセント結合処理モジュールの構築を念頭に置き，高品質なアクセントラベリングが施されたコーパスを構築すると共に，CRF を用いた統計的アクセント結合処理を実装した。従来の規則から得られる知見をコーパスベースの統計的手法に組み入れることで，従来手法や単純な統計的処理に比して精度に大きく改善が見られた。複合名詞の問題など未解決の問題があるものの，高い推定精度を得ることができた。

文　　献

[1] 匂坂芳典, 佐藤大和: “日本語単語連鎖のアクセント規則”, 電子通信学会論文誌, vol.J66-D, no.7, pp.847–856, 1983.

[2] N. Minematsu, R. Kita, and K. Hirose, “Automatic estimation of accentual attribute values of words for accent sandhi rules of Japanese text-to-speech conversion,” Trans. IEICE, vol.E86-D, no.3, pp.550–557, 2003.

[3] 長野徹, 森信介, 西村雅史: “N-gram モデルを用いた音声合成のための読みおよびアクセントの同時推定”, 情報処理学会論文誌, vol.47, no.6, pp.1793–1801, 2006.

[4] 国立国語研究所報告 124「日本語話し言葉コーパスの構築法」, 独立行政法人国立国語研究所, 2006.

[5] J. Lafferty, A. McCallum, F. Pereira: “Conditional Random Fields: Probablistic Models for Segmenting and Labeling Sequence Data”, Proceedings of the 18th International Conference on Machine Learning, pp.282-289, 2001.

[6] 工藤 拓: “CRF++: Yet Another CRF toolkit” http://chasen.org/˜taku/software/CRF++/