条件付き確率場を用いた日本語東京方言のアクセント結合自動推定

鈴木　雅之[†a)]　黒岩　龍[†*b)]　印南　圭祐[†**c)]　小林　俊平[†***d)]
清水　信哉[†****e)]　峯松　信明[†f)]　広瀬　啓吉[†g)]

Accent Sandhi Estimation of Tokyo Dialect of Japanese Using Conditional Random Fields

Masayuki SUZUKI[†a)], Ryo KUROIWA[†*b)], Keisuke INNAMI[†**c)], Shumpei KOBAYASHI[†***d)], Shinya SHIMIZU[†****e)], Nobuaki MINEMATSU[†f)], and Keikichi HIROSE[†g)]

あらまし　日本語テキスト音声合成において，任意の入力テキストに対し正しいアクセントを推定することは，自然な合成音声を得るために不可欠である．日本語は，単語が文中で発声されると，アクセントが前後の文脈に応じて変化する，アクセント結合と呼ばれる現象が発生する．本研究では，この日本語のアクセント結合を統計的に自動推定する課題に取り組む．まず本研究の遂行に必要な，文発声時のアクセント情報がラベル付けされた文章データベースを作成した．ここでは 6334 文の日本語文セットを対象に，日本語東京方言話者の作業者一名が，アクセント句境界，文中の単語アクセント型のラベリングを行った．そしてこのデータベースを利用し，条件付き確率場を用いた日本語東京方言のアクセント句境界及び文中の単語アクセント型推定手法を提案する．アクセント句単位でアクセント結合自動推定の正答率を調べたところ，規則処理 (87.48%) と比較して，提案手法 (94.66%) はより高精度にアクセント結合を推定できることが示された．更に規則処理によるアクセント結合処理を用いた合成音声と，提案によるアクセント結合処理を用いた合成音声とを，聴取実験により比較したところ，提案手法は合成音声の自然性を有意に向上させられることが分かった．

キーワード　日本語テキスト音声合成，アクセント結合，アクセント句境界推定，アクセント型推定，条件付き確率場

1. まえがき

スマートフォンの音声対話システム，カーナビの音声案内システム，視覚障がい者のためのテキスト読み上げシステムなど，日本語のテキストからの音声合成 (Text To Speech; TTS) を利用した様々なアプリケーションが登場してきている．しかし現在の日本語 TTS による合成音声は，特に韻律的特徴において不自然な部分があるため，これを解決すべく研究が進められている．

広く利用されている日本語 TTS システムの処理の概要を図 1 に示す [1]．システムに任意のテキストが入力されると，まず辞書を参照しながら形態素解析を行い，その読みを推定する．例えば，「200 m」というテキストから，「ニヒャクメートル」という読みを出力する．次に，推定した読みに対し，適切なアクセント情報を付与する．例えば東京方言では，アクセントは各モーラの高低で記述でき，図 1 の中段に示す高低パターンで発声するのが正しいため，この情報を推定して出力する．最後に，これらの読みやアクセント等の情報を用い，音声波形を出力する．

ここで日本語は，単語を単独で発声した場合と文中で発声した場合とでアクセントが変化する，アクセン

† 東京大学，東京都
The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
* 現在，(株) NTT データ
** 現在，富士通株式会社
*** 現在，(株) 野村総合研究所
**** 現在，マッキンゼー・アンド・カンパニー・インク・ジャパン
a) E-mail: suzuki@gavo.t.u-tokyo.ac.jp
b) E-mail: kuroiwar@nttdata.co.jp
c) E-mail: yrpnov20@gmail.com
d) E-mail: shumpei.0601@gmail.com
e) E-mail: shinya@fgresearch.jp
f) E-mail: mine@gavo.t.u-tokyo.ac.jp
g) E-mail: hirose@gavo.t.u-tokyo.ac.jp

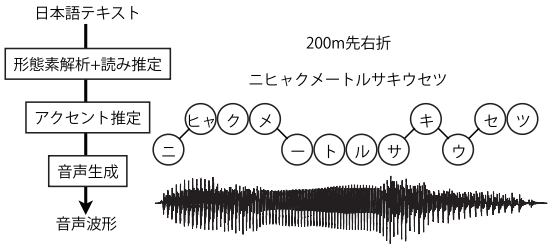

図 1　日本語 text-to-speech システムの処理の概要
Fig. 1　An overview of Japanese text-to-speech systems.

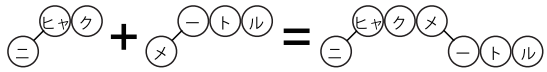

図 2　日本語のアクセント結合の例
Fig. 2　An example of Japanese accent sandhi.

ト結合と呼ばれる現象が頻繁に発生する．例えば「ニヒャクメートル」の例では，図 2 のようなアクセント結合が発生している．アクセント結合処理は，日本語母国話者は無意識的に正しく行っているが，単語のアクセントが孤立発声時からどのように変化するのかを，規則として完全に明文化するのは難しい．推定された文中アクセントが不適切であれば，当然ながら TTS による合成音声の自然性が低下してしまうため，精度の高いアクセント結合推定器の実現が望まれる．

本研究の目的は，精度の高い日本語東京方言アクセント結合推定器を実現し，TTS の性能を向上させることである．我々はまず，日本語 6334 文に対して，それらを文として発声した場合のアクセント句境界と各単語のアクセントの情報をラベル付けしたデータベースを作成した．そしてそのデータベースを利用した，条件付き確率場 (Conditional Random Fields; CRF) による統計的なアクセント結合自動推定法を提案する．提案手法のアクセント句単位での正答率を調べたところ，94.66%となり，従来広く利用されている規則処理を用いた場合の 87.48%と比較して，高精度にアクセント結合を推定できることが分かった．更に，オープンソースの日本語 TTS システムとして広く利用されている Open JTalk [2] のアクセント結合処理を提案手法に置き換えたところ，合成音声の自然性が有意に高くなることが分かった．

2.　日本語東京方言アクセントの定義

アクセントと呼ばれる現象は言語により異なる音響的特徴によって実現されているが，日本語は，モーラ単位の音の高さの変化による高低アクセントをもつ．ここでモーラとは，日本語発声の基本単位であり，拗音を除くとカナ一文字がモーラ一つに対応する．モーラには，母音，子音+母音，子音+半母音（拗音）+母音で構成される一般モーラと，長母音，促音，撥音で構成される特殊モーラとがある．

一つ以上の形態素が連なって作られるアクセント的まとまりのことをアクセント句と呼ぶ．アクセント句の境界は，必ず形態素の境界でもある．アクセント句と文節は，しばしば一致する．東京方言においては，アクセント句内に，たかだか一つ，音の高さが下降する箇所がある．この下降が起こる直前のモーラのことをアクセント核と呼ぶ．各アクセント句は，アクセント核の位置が N モーラ目にあるものを N 型，アクセント核がないものを 0 型と，アクセント型を使って分類できる．1 型のアクセント句以外では，音の高さの下降に加えて，アクセント句の 1 モーラ目から 2 モーラ目にかけて音の高さの上昇が発生する．例えば，図 1 のようなアクセントは，「ニヒャクメートル/サキ/ウセツ」のように三つのアクセント句で，一つ目のアクセント句には先頭から 4 モーラ目の「メ」にアクセント核がある 4 型，二つ目と三つ目のアクセント句は，アクセント核がない 0 型ということになる．なお，アクセントは話速等によっても変化する場合があり，例えば先の例では「ニヒャークメートルサキ/ウセツ」と二つのアクセント句に分かれ，二つとも 0 型であるとしても，東京方言として許容されるアクセントである．

以上のようなアクセント句とアクセント型の定義は，実際には一つのアクセント句内に二つ以上のアクセント核が生じる副次アクセントなど，日本語東京方言のアクセントに関する現象を厳密に網羅する定義ではないが，工学的な利便性から，本論文ではこの定義を採用する．この定義を用いると，本研究の目的である日本語東京方言のアクセント結合推定は，1) 形態素列を入力としてアクセント句境界を推定するタスク，2) アクセント句のアクセント型を推定するタスク，の二つに分割できる．

3.　規則に基づくアクセント結合推定

3.1　規則に基づくアクセント句境界推定

形態素列から，規則に基づいてアクセント句境界を推定する手法としては，例えば Open JTalk に実装されている規則がある．Open JTalk version 1.05 では，ある二つの形態素に挟まれた形態素境界がアクセント

表 1 Open JTalk version 1.05 のアクセント句境界推定規則
Table 1 Rules of estimating accent phrase boundaries used in Open JTalk version 1.05.

前の形態素	後の形態素	境界
形容詞	名詞	あり
形容動詞語幹	名詞	あり
動詞	名詞	あり
接尾辞	名詞	あり
動詞	形容詞	あり
付属語	自立語	あり
どちらかが 副詞/名詞，副詞可能		あり
どちらかが 接続詞/連体詞/記号		あり
その他		なし

句境界か否かを推定するために，表 1 のような規則が実装されている [2]．すなわち，形態素境界前後の形態素の品詞情報のみに依存して，そこにアクセント句境界があるか否かを決定している．これは非常に簡単に実装できるものの，精度には限界がある．

3.2 規則に基づくアクセント型推定

アクセント句から規則に基づいてアクセント型を推定する手法としては，匂坂らによるアクセント結合規則（匂坂規則）が広く知られている [3]．匂坂規則には，二つの単語がアクセント結合したときにどのようなアクセント型になるかが定められており，その規則をアクセント句内で左から巡回的に適用していくことで，最終的にアクセント句のアクセント型を一意に決定する．それぞれの規則は，文節の構成要素（品詞など），単独発声した場合のアクセント型，モーラ数，アクセント結合様式といった情報から，二つの単語が結合したときのアクセント型を決める．例えば，「東京大学」というアクセント句は，「読み：トーキョー/品詞：名詞/モーラ数：4/単独発声アクセント型：0 型」と「読み：ダイガク/品詞：名詞/モーラ数：4/単独発声アクセント型：0 型/アクセント結合様式：C2」といった情報を形態素解析により得て，「名詞連続のアクセント結合で，後続名詞のアクセント結合様式が C2 の場合，前の形態素のモーラ数＋1 モーラ目にアクセント核が生じる」という規則を適用し，アクセント型が 5 型であると決定する．

なお匂坂規則では，数詞や助数詞のアクセント結合の結果が不自然になる場合が多いので，これに関しては別の規則を同時に利用することが多い．数詞に特化したアクセント結合規則としては，例えば宮崎らによる規則（宮崎規則）がある [4]．また Open JTalk version 1.05 には，匂坂規則と，独自の数詞に関するアクセント結合規則を組み合わせたものが実装されている．

4. 文中アクセントラベルデータベース

これまで，上記のような規則ベースによるアクセント句境界推定・アクセント型推定が研究されてきているが，それらの精度は必ずしも高くない．我々の研究グループでも，規則を更に改良する研究を行ってきたが，大幅な精度向上は実現できなかった [5]．

自然言語処理は歴史的に，規則ベースの手法から，大量のラベル付きデータベースを利用した統計ベースの手法に移行することで，大幅な精度向上を実現してきた [6]．そのためアクセント結合推定も，規則ベースでなく，統計ベースで行うことで，精度向上が実現できると予想される．実際既に，アクセント推定タスクにおいて，統計ベースの手法の有効性が報告されている [7]〜[9]．

しかしこれまで，一般公開されている大規模な文中アクセントをラベル付けしたデータベースが存在しておらず，これが統計ベースのアクセント結合推定の研究があまり行われていない一つの原因となっていた．そこで我々はまず，大量の日本語文章に対して，それらを文として東京方言で発声した場合のアクセント句境界と各アクセント句のアクセント型の情報をラベルとして付与したデータベースを構築した．以降，このデータベースのことを，「文中アクセントラベルデータベース」と呼ぶ．以下，我々が構築した文中アクセントラベルデータベースの仕様について簡単に述べる．詳細に関しては，[10] を参照されたい．

ラベリングの対象となる文は，JNAS [11] や S-JNAS [12] で使用されている文から選ばれた 6334 文である．これらを，UniDic [13] で利用されている短単位を利用して形態素解析し，手動で読みを修正したものに対してアクセント句境界と各アクセント句のアクセント型のラベリングを行った．この際アクセント句内のアクセント核の出現数は制限していない．副次アクセントとして二つ目以上のアクセント核を付けるべきか，複数のアクセント句に分けるべきかを判断する基準には，句頭の音の高さの上昇があるべきか否かを用いた．ただし，一つの形態素の中には，たかだか一つのアクセント核しか存在しないと制限した．

ラベリング作業は，方言や個人によるアクセント感覚の違いの影響を取り除くため，音感に優れた東京出身東京方言話者の作業者 1 名のみが，音声データを

用いず文字テキストのみを参照して行うものとした．アクセント句は，話速によっても変化するため，約 7 モーラ/秒の速さで自然に読んだ場合を想定させた．ラベリングの誤りを防ぐため，別の東京出身東京方言話者の作業者がチェックを行い，不自然と思われる箇所については，先の作業者が再度ラベリングを行った．このようにラベリングを行ったため，本データベースは，厳密には日本語東京方言のデータベースではなく，ラベリングを行った作業者のアクセント感覚がラベル化されたデータベースとなっている．

なお構築した文中アクセントラベルデータベースは，JNAS 若しくは S-JNAS 購入者に無償配布している．第一著者若しくは第六著者に連絡されたい．

5. CRF を用いたアクセント結合推定

5.1 条件付き確率場

本論文では，アクセント結合推定に条件付き確率場 (Conditional Random Field; CRF) を利用する手法を提案する．そこでまず，CRF に関して，簡単な説明を行う [14]．

CRF は，系列ラベリング問題を解くのに利用できる識別モデルである．本論文では，形態素系列に対し，各形態素のアクセント情報を表すラベルを推定するのに CRF を利用する．観測データ系列（形態素の系列）を $\boldsymbol{x}$，それに対するラベル系列を y，ラベルのとりうる値の集合を Y として，CRF は，ラベル系列の事後確率を，下式でモデル化する．

$$p(y|\boldsymbol{x}) = \frac{\exp \sum_{f=1}^{F} w_f \phi_f(\boldsymbol{x}, y)}{\sum_{y' \in Y} \exp \sum_{f=1}^{F} w_f \phi_f(\boldsymbol{x}, y')} \qquad (1)$$

ここで，$\{\phi_f(\boldsymbol{x}, y)\}_{f=1 \cdots F}$ は素性関数と呼ばれ，本論文では 0/1 のどちらかの値しかとらないと限定する．F は素性の総数である．素性関数は観測素性に基づくもの，遷移素性に基づくものに分けることができる．観測素性に基づく素性関数は，特定のラベルかつそれに対応する観測データが何らかの特徴を満たす場合のみに 1，そうでなければ 0 となる関数である．以降，観測データのある特徴と全てのラベルを用いた観測素性を利用することを，「CRF にある特徴を用いる」と書く．ただし，この特徴は，離散的で有限個の値のみをとる特徴とする（例えば形態素の品詞）．この際，観測データにその特徴が定義できない場合は，そのことを表す undefined ラベルを値として利用する．CRF にある特徴を用いると，観測素性として，（特徴がとりうる値の数）×（ラベルがとりうる数）の素性関数が追加されることになる．遷移素性としては，本論文では，ラベル系列が特定の bigram となった場合に 1 となり，そうでなければ 0 となる素性関数のみを用いる．遷移素性を利用するかしないかが，多クラスロジスティック回帰と CRF の，ラベル事後確率の定義における唯一の違いである．$\{w_f\}_{f=1 \cdots F}$ は，各素性に対する重みであり，これが CRF のモデルパラメータである．

5.2 CRF を用いたアクセント句境界推定

形態素列からアクセント句境界を推定するタスクは，形態素ごとに，当該形態素の直前にアクセント句境界があるか否かを推定するタスクとして定式化する．具体的には，$\boldsymbol{x}$ を一文分の形態素系列，y を当該形態素の直前にアクセント句境界が存在するかしないかの 0/1 ラベル系列とし，$p(y|\boldsymbol{x})$ を CRF でモデル化する [15]．そして，この事後確率が最も大きくなる y を，推定結果とする．

アクセント句境界推定のために用いる CRF で利用する特徴を，表 2 にまとめた．まず第一に，前後の形態素の品詞情報を利用する．これは，Open JTalk のアクセント句境界推定規則（表 1）でも利用されてお

表 2 CRF を用いたアクセント句境界推定で用いる特徴
Table 2 Feature types for CRF-based accent phrase boundary estimation.

以下は当該形態素の二つ前から二つ後の形態素五つ分の特徴それぞれを，全て当該形態素の特徴として用いる．	
a	品詞
b	書字形，発音形，活用型の組
c	活用型
d	活用形
e	語種
f	語頭変化結合型
g	単独発声アクセント型
h	アクセント修飾価
i	直前に文節区切りがあると推定されたかの 0/1
j	当該形態素のモーラ数と二つ前から二つ後の形態素のモーラ数の組五つ分
k	当該形態素の単独発声アクセント型と二つ前から二つ後の形態素のアクセント結合型の組五つ分
l	常に 1 となる特徴（バイアス項）
m	アクセント句境界 0/1 ラベルの bigram（遷移素性）
以下は学習データを 5 等分するしきい値で 1/2/3/4/5 に離散化	
n	前の名詞と当該名詞の bigram の出現頻度
o	前の名詞と当該名詞の bigram の出現頻度を，前の名詞の unigram 出現頻度で割った値
p	前の名詞と当該名詞の bigram の出現頻度を，当該名詞の unigram 出現頻度で割った値
q	前の名詞と当該名詞の bigram の出現頻度を，前の名詞と当該名詞の unigram 出現頻度で割った値

り，アクセント句境界推定に有効だと考えられる．他にも，活用型などといった形態素の属性も，特徴として利用した．なお，各属性は，UniDic に登録されているもののみを利用しているため，属性の定義については UniDic を参照されたい [13]．これらの特徴に加え，アクセント句境界と文節境界は一致することが多いため，直前に文節境界があると推定されたか否かの 0/1 も，特徴として利用した．

ここで，アクセント句境界推定は，特に名詞と名詞が連続する際に，その間に境界があるのか否かを判別することが難しい．例えば「東京大学工学部」は，「東京大学/工学部」と区切るのは適切だが，「東京/大学工学部」は不自然である．この問題に対処するため，あらかじめ名詞連続に関する形態素 N-gram を学習しておき，それに基づくスコアをアクセント句境界推定の特徴として利用することにした（表 2 の n，o，p，q）．これにより，比較的連続して出現しやすい「東京」と「大学」の間にはアクセント句境界がなく，比較的連続しにくい「大学」「工学部」の間にはアクセント句境界がある，といったように，適切にアクセント句境界が推定されることが期待される．なお，N-gram に関する実数値スコアを離散値のみを取り扱う CRF の特徴として用いるために，学習データを 5 等分するようにスコアのしきい値を決定しておき，それに基づきスコアを 1/2/3/4/5 の 5 値のいずれかをとる特徴量とした．

5.3 CRF を用いたアクセント型推定

アクセント句からアクセント型を推定するタスクは，アクセント句内の各形態素を単独で発声した場合のアクセント型が，文中でどのように変化するのかを表す相対変化ラベルを推定するタスクとして定式化する．

まず相対変化ラベルについて説明する [10]．文中での形態素のアクセント核位置は，あらゆる位置にアクセント核が生じ得るわけではなく，ほとんどの場合，ある特定のアクセント核位置の変化パターン（相対変化パターン）をとる．具体的には，以下の ***V*** から ***P*** の 7 パターンのいずれかとなる．

- ***V***anish：単独発声時の核がなくなる
- ***R***emain：単独発声時の核がそのまま残る
- ***N***ever：単独発声時もアクセント結合後も無核
- ***B***efore：単独発声時の核の一つ前が核になる
- ***L***ast：末尾のモーラが核になる
- ***F***irst：1 モーラ目が核になる
- ***P***enultimate：末尾の一つ前が核になる

ただし，複数の条件にあてはまる場合は，先に書いた方のパターンを採用させる．また，数は非常に少ないものの上記のいずれにもあてはまらない場合は，もとのアクセント核位置（0 型の場合は 0）から何モーラ後ろに核が移動したかの数字 $(\mathbf{1}, \mathbf{2}\cdots)$ を，相対変化ラベルとして用いる．以上のような相対変化ラベルを利用すると，形態素ごとに上記のラベルのいずれになるかを識別するだけで，効率的にアクセント句のアクセント型を決定することができる．これを，$\boldsymbol{x}$ をアクセント句内の形態素系列，y をそれに対応する相対変化ラベル系列として，$p(y|\boldsymbol{x})$ を CRF でモデル化することで実現する．

アクセント相対変化ラベル推定のために用いる CRF で利用する特徴を，表 3 にまとめた．まず第一に，句坂規則でも利用されている，品詞，単独発声アクセント型，モーラ数，アクセント結合様式などといった情報が有効だと考えられるため，これを特徴として用いる．他にも，様々な特徴を利用している，以下，特筆すべき特徴に関して詳しく説明を行う．

修正された単独発声アクセント型の第一候補　UniDic に品詞の属性として登録されているアクセント修飾型は，特定の活用をとる場合に，基本形の単独発声アクセント核位置がどう変化するかを表している．そこで，アクセント修飾型に基づいて単独発声アクセント型を修正したものを特徴として利用することにした．また，UniDic には複数の単独発声アクセント型候補が記述されている場合があるので，その第一候補のみを利用した．

規則に基づくアクセント相対変化ラベル　単純な自立語と付属語の 2 形態素からなるアクセント句では，ほとんどの場合，句坂規則で正しいアクセント型を推定することができる．そこで，句坂規則・宮崎規則から推定した相対変化ラベルも，CRF の特徴として利用することにした．

h の種類ラベル　句坂規則では，場合分けとして，

- アクセント核があるか否か
- 末尾に核があるか否か
- 末尾の一つ前に核があるか否か
- その他

が利用される．そこで，この 4 パターンを種類ラベルとして特徴に利用することで，句坂規則による知見を文中アクセント型推定に導入することができると考えられる．なお，「末尾の一つ前に核がある」場合に関しては，末尾 2 モーラに重音節を含むか否かが規則に関係するため，末尾 2 モーラそのものを種類ラベルと

表 3　CRF を用いた相対変化ラベル推定で用いる特徴
Table 3　Feature types for CRF-based estimation of labels for relative accent sandhi.

以下は当該形態素の二つ前から二つ後の形態素五つ分の特徴それぞれを，全て当該形態素の特徴として用いる．	
a	品詞
b	単独発声アクセント型
c	モーラ数
d	動詞に対するアクセント結合様式
e	形容詞に対するアクセント結合様式
f	名詞に対するアクセント結合様式
g	アクセント修飾型
h	修正された単独発声アクセント型の第一候補
i	規則に基づくアクセント相対変化ラベル
j	h の種類ラベル
k	書字形
l	発音形
m	活用型
n	活用形
o	語彙素
p	語種
q	語頭変化結合型
r	アクセント句の一つ目の形態素か否かの 0/1
s	アクセント句内の形態素数
t	IREX の定義に基づく固有表現タグ推定値 [16]
u	2 モーラ以下か否かの 0/1
v	2 モーラ以下か否かの 0/1 と，語種の組
w	重音節を含むか否かの 0/1
x	先頭のモーラ
y	先頭から二つめのモーラ
z	アクセント核の一つ前のモーラ
A	アクセント核のモーラ
B	アクセント核の一つ後のモーラ
C	末尾の一つ前のモーラ
D	末尾のモーラ
E	規則から推定したアクセント相対変化ラベルと，当該形態素と一つ前の形態素の品詞の組
F	当該形態素の h と当該形態素を除く二つ前から二つ後の形態素のアクセント結合型の組四つ分
G	当該形態素のアクセント結合型と当該形態素を除く二つ前から二つ後の形態素の h の組四つ分
H	当該形態素の品詞，h と当該形態素を除く二つ前から二つ後の形態素の $[d\|e\|f]$ の組計 $3 \times 4 = 12$ つ分
I	当該形態素の $[d\|e\|f]$ と当該形態素を除く二つ前から二つ後の形態素の品詞，h の組計 $3 \times 4 = 12$ つ分
J	常に 1 となる特徴（バイアス項）
K	相対アクセント変化ラベルの bigram（遷移素性）
以下は数詞/助数詞を適切に取り扱うための特徴	
L	当該形態素の語頭変化結合型と当該形態素から 1 or 2 つ後の形態素の助数詞タイプ二つ分
M	当該形態素が数詞か否かの 0/1 と当該形態素から 1 or 2 つ後の形態素の助数詞タイプ二つ分
N	当該形態素の語頭変化結合型と当該形態素から 1 or 2 つ後の形態素が助数詞か否かの 0/1 二つ分
O	当該形態素が数詞か否かの 0/1 と当該形態素から 1 or 2 つ後の形態素が助数詞か否かの 0/1 二つ分
P	当該形態素の助数詞タイプと当該形態素から 1 or 2 つ前の形態素の語頭変化結合型二つ分
Q	当該形態素の助数詞タイプと当該形態素から 1 or 2 つ前の形態素が数詞か否かの 0/1 二つ分
R	当該形態素の助数詞か否かの 0/1 と当該形態素から 1 or 2 つ前の形態素の語頭変化結合型二つ分
S	当該形態素の助数詞か否かの 0/1 と当該形態素から 1 or 2 つ前の形態素が数詞か否かの 0/1 二つ分

表 4　助数詞のタイプ分類
Table 4　A classification table for counter suffixes.

a	個，位，時，分（ふん），時間，歳，羽，通り，斤，層，アール，センチ，キロ，ドル，度（ど：温度，角度），階，球，巡，乗，週，人前，敗，着（到着），度目，代目，貫目，幕目，日目，球目，丁目，畳，ヶ月
b	問，台，軒，票，町，艘，代，枚，名，面，本，枚，丁
c	升
d	年（ねん），段（階段），番
e	貫，版，銭，回，点，巻
f	尺，着（衣服），角
g	円
h	曲，石（こく），匹，冊，足，拍，脚，局，発
i	合
j	度（ど：回数）
k	人
l	月（がつ），日（にち）
m	寸

して用いた [10]．

2 モーラ以下か否か/重音節を含むか否かの 0/1　外来語のアクセント型は，2 モーラ以下か 3 モーラ以上か否かや，重音節を含むか否かによってアクセント核位置が変化しやすいことが知られている．そこでこれらの特徴を利用することで，外来語の相対変化ラベルを適切に推定しやすくなると考えられる [17]．

数詞/助数詞を適切に取り扱うための特徴　宮崎らの研究により，助数詞のタイプによって，数詞，助数詞のアクセント核位置が変化することが知られている．助数詞タイプの表を表 4 に示す．これを特徴として利用することで，数詞の相対変化ラベルを適切に推定しやすくなると考えられる [17]．

6. 実　　験

6.1 アクセント句境界推定

提案手法である CRF を用いたアクセント句境界推定を，規則ベースの手法と比較する実験を行った．データベースには，先述のとおりに構築した日本語東京方言文中アクセントラベルデータベース 6334 文を用いた．まず最初に，MeCab version 0.993 [18]，CaboCha version 0.62 [19]，UniDic version 1.3.12 [13] をもとに学習された MeCab/CaboCha の付属のモデルを利用して，形態素解析，読み推定，文節境界推定，Information Retrieval and Extraction Exercise (IREX) の定義に基づく固有表現タグ推定 [16] を行った．これらの推定結果は，特に読み推定において多くの誤りを

含んでいる．読み推定が誤っている場合，たとえ誤った発音が正しいと仮定した上での正解アクセントが推定できたとしても，最終的な TTS の結果は不自然となってしまうと考えられる．そこで今回はアクセント結合推定のみに注目するため，形態素解析誤りと読み誤りを含む文をデータベースから全て削除することにした．この処理により，データベースは 4785 文に減少する．次にこの文セットを 3786 文の学習データ，999 文の評価データにランダムに分割した．学習データは 66048 形態素，評価データは 17801 形態素を含んでいる．そのうち直前にアクセント句境界をもつ形態素は，学習データには 25542，評価データには 7641 ある．

CRF の特徴には，先述のとおり表 2 の特徴を用いた．特徴の抽出に利用する名詞連続の形態素 bigram は，2012 年 4 月 10 日における日本語版 wikipedia 全記事のダンプ結果を，WP2TXT version 0.1.0 [20] を利用してテキスト化し，それを MeCab + UniDic で形態素解析したものから学習した．CRF の実装には，CRF++ version 0.57 [21] を利用した．正則化パラメータ以外に関しては，CRF++ のデフォルトの設定をそのまま利用した．すなわち，CRF のモデルパラメータの学習は，二次の正則化項付きの目的関数を L-BFGS アルゴリズムで最適化することで行った．正則化パラメータは，学習データを利用して 3-fold クロスバリデーションによるグリッドサーチを行うことで決定した．具体的には，crf_learn の -c オプションに設定する値を 10，5，1，0.5，0.1，0.05，0.01 と変化させ，最も平均精度（F 値）が高くなる 0.1 を採用した．この値を用いて，学習データ全てを利用して CRF のパラメータを学習し，評価データの全ての形態素の直前にアクセント句境界が存在するかしないかの 0/1 ラベルを推定した．

また，規則ベース処理として，Open JTalk で利用されている規則（表 1）を利用した．この際，表 1 に書かれている処理は，Open JTalk で利用されている NAIST Japanese Dictionary [22] の品詞体系によるものであるため，これを UniDic 体系に適切に読み替えて実装した．

規則と CRF それぞれの，正答数，脱落誤り数，挿入誤り数，適合率，再現率，F 値を表 5 に示す．提案手法である CRF を用いた手法は，規則を用いた手法と比較して，適合率でも再現率でも精度が向上しており，特に適合率は大幅に精度が向上している．F 値では約 5 ポイント精度が上がっており，提案手法の有効性が示された．

表 5　アクセント句境界推定の実験結果
Table 5　Results of accent phrase boundary estimation.

	正答数	脱落	挿入	適合率	再現率	F 値
規則	6804	837	871	89.1%	88.7%	88.9
CRF	6915	726	182	97.4%	90.5%	**93.8**

表 6　名詞連続部分のアクセント句境界推定の実験結果
Table 6　Results for the case of compound nouns comprising two words.

	正答数	脱落	挿入	適合率	再現率	F 値
規則	0	606	0	-	-	-
CRF	395	211	60	65.2%	88.7%	**74.5**
$\mathrm{CRF}_{\mathrm{w/o}N\text{-gram}}$	380	226	65	62.7%	85.4%	72.3

次に，名詞連続部分のみに注目した結果を表 6 に示した．評価データには，名詞に挟まれた形態素境界が 1760 あり，そのうち 606 にはアクセント句境界があるとラベル付けされている．参考に，名詞連続 N-gram に関する特徴を取り除いて CRF でアクセント句境界を推定した結果も示している ($\mathrm{CRF}_{\mathrm{w/o}N\text{-gram}}$)．結果まず，規則を用いた場合，名詞と名詞の間には必ずアクセント句境界がないと判定してしまうため，正答数が 0 になってしまうことが分かる．一方 CRF を用いると，脱落誤り数は多いものの，ある程度の精度でアクセント句境界が検出できていることが分かる．CRF と $\mathrm{CRF}_{\mathrm{w/o}N\text{-gram}}$ を比べると，CRF の方が F 値が高いことから，名詞連続 N-gram を利用することの効果があることが分かる．しかしながら効果は限定的であり，今後の研究による精度向上が望まれる．具体的には，名詞連続のアクセント句境界は形態素間の係り受け，格，意味等に依存して決まると考えられるため，こういった情報を推定に利用することが考えられる．

6.2　アクセント型推定

提案手法である CRF を用いたアクセント型推定を，匂坂・宮崎規則ベースの手法と比較する実験を行った．データベースには，先のアクセント句境界推定と同様の処理を行い，3786 文の学習データ，999 文の評価データを用意した．アクセント句境界としては，データベースに用意付与されている正解アクセント句境界を利用するものと，先の実験で推定したアクセント句境界（規則ベースのものと，CRF ベースのもの），合計 3 通りを用いて実験を行った．正解アクセント句境界を利用する場合は，学習データに 25542，評価データに 7641，アクセント句がある．同様に規則から推定

表 7　アクセント型推定の実験結果
Table 7　Results of accent sandhi estimation.

句推定	型推定	正解数	総数	正答率
正解	規則	6900	7641	90.30%
正解	CRF	7420	7641	**97.11%**
規則	規則	6714	7675	87.48%
規則	CRF	7251	7675	**94.48%**
CRF	規則	6218	7097	87.61%
CRF	CRF	6718	7097	**94.66%**

したアクセント句境界を利用する場合は学習データに 28612，評価データに 7675，CRF で推定したアクセント句境界を利用する場合は学習データに 26076，評価データに 7097，アクセント句がある．規則や CRF で推定したアクセント句境界と正解の境界が一致しなかった場合のアクセント型正解ラベルは，正しいアクセント核位置をもとに決定した．そのため，各アクセント句単独で見ると不自然な正解ラベルが付与されている場合もあるが，文全体としては，自然なアクセントになるように正解ラベルが付与されることになる．このように定義したデータには，副次アクセントやアクセント句境界推定の誤りにより，一つのアクセント句内に二つ以上の核が存在する場合がある．この場合は，先に出現した核を主なアクセント核とし，二つ目以降のアクセント核を集計から除外した．同様に，アクセント核位置の推定結果に二つ以上の核があると推定される場合にも，先に出現した核を主なアクセント核とし，二つ目以降のものは除外した．

CRF の特徴には，先述のとおり表 3 の特徴を用いた．実装には，アクセント句境界推定と同様 CRF++ を利用した．CRF の正則化パラメータは，3-fold クロスバリデーションにより，crf_learn の -c オプションに設定する値を 1.2，1.0，0.8，0.6，0.4 と変化させ，最も平均正解率が高くなる 0.8 を採用した．この値を用いて，学習データ全てを利用して CRF のパラメータを学習し，これを用いて評価データのアクセント相対変化ラベルを推定し，それをもとにアクセント型を決定した．

結果を表 7 に示す．提案手法である CRF を用いたアクセント型推定は，規則を用いたものと比べて，どのようなアクセント句境界推定結果を用いても，精度が向上することが分かった．アクセント句境界及びアクセント型両方に規則を用いた場合は 87.48%，両方に CRF を用いた場合は 94.66%となることから，提案手法により大幅な精度が実現できたことが分かる．アクセント句境界は正解を与えた場合には，CRF を用いた提案手法は 97%以上という高い正答率を示した．残りの約 3%含まれる誤りのうち，先頭から 20 個の誤りの例を表 8 に示す．これを見ると，許容できない CRF 推定誤りは 3%よりも少ないと予想される．

表 8　アクセント句境界に正解を与え CRF でアクセント型を推定した場合の誤りの例
Table 8　Examples of false accent sandhi estimation when correct accent phrase boundaries are given.

アクセント句	正解	CRF	アクセント句	正解	CRF
おらず	2 型	1 型	火の気が	0 型	1 型
最高だなあ	6 型	0 型	シンバスタチン	5 型	4 型
低く	2 型	1 型	赤い色素も	5 型	4 型
だれも	0 型	1 型	同国や	1 型	0 型
来年度版からの	9 型	5 型	軍属	0 型	1 型
五年計画で	4 型	0 型	ともに	1 型	0 型
ひどく	2 型	1 型	原子力	0 型	3 型
いない	0 型	2 型	ものの	0 型	2 型
景気回復局面で	8 型	9 型	強く	2 型	1 型
今月末にも	4 型	3 型	さすがに	0 型	3 型

6.3　TTS の性能評価

最後に，提案手法により TTS による音声合成の自然性がどの程度向上するのかを聴取実験を通して検証した．アクセント句境界推定実験，アクセント型推定実験で用いた正解は，文中アクセントラベルデータベースにある正解であり，これはラベリングを依頼した話者のアクセント感覚に合致しているか否かを評価基準としている．しかし，**2.1** でも示したように，句境界や型が上記正解以外の値をとった場合でも，それが必ずしも誤りとはならない．そこで本節では，規則処理と提案手法とによって得られたアクセント句境界，アクセント型の情報を使って二種類の合成音声を作成し，両者を比較することで，提案手法の有効性を検証する．具体的には，Open JTalk に実装されている規則ベースで推定されるアクセントラベルを用いた TTS の出力と，そのアクセントラベルを提案手法による CRF を用いたアクセント結合推定結果で置き換えて音声合成した出力とを，一対比較法を用いて自然性を比較する．

まず，TTS システムとして Open JTalk version 1.05 [2]，hts_engine API version 1.06 [23]，MMDAgent 付属の HTS Voice “Mei (Normal)” version 1.1 [24] による HMM 音声合成システムを利用する．サンプリング周波数は 48000 Hz，フレームピリオドは 240 point，all-pass constant は 0.55 とした．Open JTalk では，文を NAIST Japanese Dictionary を利用して形態素解析や読み推定した結果を利用するた

め，我々が先の実験で利用した UniDic の分析結果と異なる読みになる場合がある．そこで，TTS の対象は，CRF を用いたアクセント結合推定で利用した評価データ 999 文のうち，読み推定の結果が一致する 873 文を抽出，更にそこからランダムに選んだ 50 文とした．Open JTalk には，規則ベースのアクセント句境界推定，アクセント型推定が実装されている．規則の内容は，辞書が異なるために異なる部分や，数詞のアクセントに関する部分でわずかな違いはあるものの，先の実験で用いた規則と基本的には同じである．提案手法のアクセント結合推定結果を用いる場合は，Open JTalk から作成された HTS Voice 用コンテクストラベルを，アクセント句境界，アクセント核位置のみを置き換え，そのコンテクストラベルを用いて HMM 音声合成した．なお，コンテクストラベルの定義による制約上，一つのアクセント句にはたかだか一つのアクセント核しかもてないため，CRF でアクセント核を推定した結果二つ以上の核が存在した場合には，先の実験と同様，二つ目以降のアクセント核を除外して利用した．また，ブレスフレーズや品詞といったアクセントに関係ないコンテクストは，Open JTalk の出力をそのまま利用した．

聴取実験は，東京に三年以上在住し日常的に東京方言を話している，日本語母語話者男女 12 名によって行った．この被験者には，ラベリングの作業者は含まれていない．先に選んだ 50 文それぞれに関して，Open JTalk の規則に基づくアクセントラベルを利用して音声合成したものと，CRF を利用して推定したアクセントのラベルに置き換えてから合成したものをヘッドフォンで聴取させ，一対比較の強制選択により，自然性がより高いものを選ばせた．この際，提示順序による影響を防ぐため，順序はランダムに入れ換えて提示した．

結果を図 3 に示す．結果，多くの文で提案法の方がより自然性が高いと判定されており，t 検定の結果，$p < 0.01$ で有意差があることが分かった．これにより，提案手法は，TTS の自然性向上に効果があるといえる．

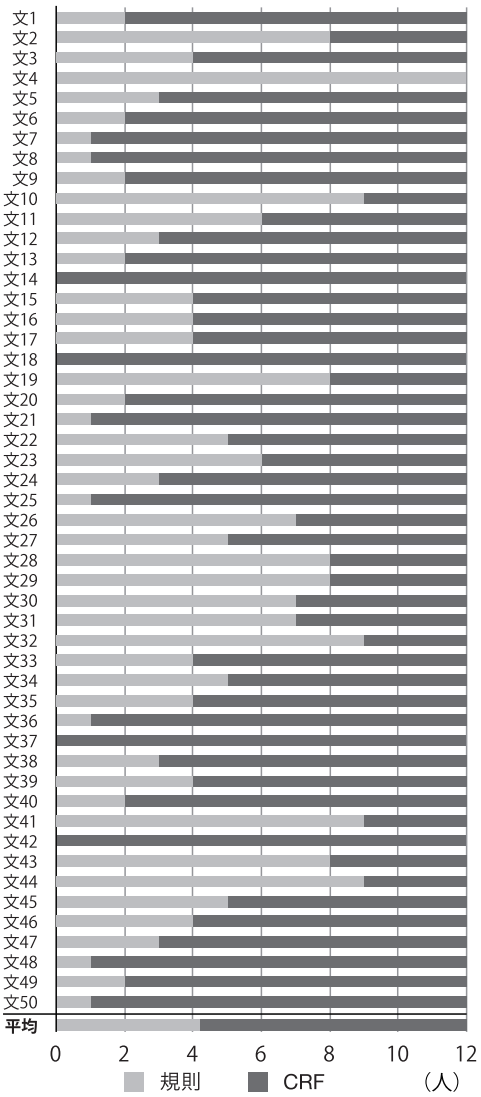

図 3　一対比較による聴取実験の結果
Fig. 3　Results of the paired comparison test.

7. 関連研究

7.1 *N*-gram による読み・アクセントの同時推定

統計ベースでアクセントを推定する手法として，長野らは，入力された文から，形態素の表層，品詞，読み，アクセントの四つ組を一つの単位とする N-gram モデルを利用し，四つを全て同時に推定する手法を提案している [7], [25]．この研究では，形態素解析処理に相当する処理とアクセント処理を同時に行った方がそれぞれを単独で行うよりも精度が高いことを示しており，この点で優れている．

しかし，モデルが前向き N-gram であるため，我々が利用している CRF のような識別モデルと比較すると，(1) 当該形態素の後ろの形態素に関する情報を使っていない (2) モデルパラメータが識別率最大化の観点で学習されない，の二点において，精度が低下してしまうことが予想される．実際，同じデータベースを用いていないため正確な比較ではないものの，長野らの

論文によると，読みが正解と一致した単語に対してアクセントが正解した割合は 92.63%となっているが，我々の提案手法（表 7 の最後の行に対応）で単語単位で正解率を集計すると，95.53%となっている．

7.2 点予測に基づく自然言語処理

近年，形態素解析等の自然言語処理の分野で，点予測に基づく処理が注目を集めている [26], [27]．点予測を利用する一番大きな利点は，部分アノテーションによる迅速な分野適応が可能なことである．本論文で取り扱った，アクセント句境界推定，文中アクセント型推定も，データベースを作成するコストが高いため，今後点予測を導入するなどして，アノテーションコストを低減していくことが必要であると考えられる．

アクセント結合推定が，点予測でも高い精度が実現できるかどうかを確かめるため，CRF で利用する特徴から，遷移素性を取り除いた実験を行ったところ，全てのタスクでほとんど差がないことが分かった．この結果は，CRF のような系列ラベリング用のモデルを用いなくても，ロジスティック回帰などの点予測で，十分にアクセント結合推定が実現できることを示唆している．これに関して，今後の研究が期待される．

8. むすび

本論文では，日本語 TTS の性能を向上させるために，CRF を用いたアクセント句境界・アクセント型推定手法を提案した．また，統計的手法でアクセント結合推定を行うために，日本語東京方言の文中アクセントラベルデータベースを構築した．提案手法を用いて，アクセント句境界推定，アクセント型推定を行ったところ，規則を用いた場合と比較して高い精度で推定が行えることが分かった．また，提案手法を用いて日本語 TTS を行った結果，規則を用いた場合と比較して，合成音声の自然性が有意に高くなることが分かった．

文　献

[1] S. Seto, M. Morita, T. Kagoshima, and M. Akamine, “Automatic rule generation for linguistic features analysis using inductive learning technique: Linguistic features analysis in TOS drive TTS system,” Proc. 5th International Conference on Spoken Language Processing (ICSLP), pp.1059–1063, 1998.

[2] Open JTalk, http://open-jtalk.sourceforge.net/

[3] 匂坂芳典，佐藤大和，“日本語単語連鎖のアクセント規則,” 信学論（D），vol.J66-D, no.7, pp.849–856, July 1983.

[4] 宮崎正弘，“日本文音声変換のための数詞読み規則,” 情処学論，vol.25, no.6, pp.1035–1043, 1984.

[5] 黒岩　龍，峯松信明，広瀬啓吉，“活用語尾に着眼した日本語アクセント結合規則の整理と高精度化,” 言語処理学会全国大会，pp.995–998, 2006.

[6] 北　研二，確率的言語モデル，言語と計算，第 4 巻，東京大学出版会，1999.

[7] 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,” 情処学論，vol.47, no.6, pp.1793–1801, 2006.

[8] 鈴木和博，山本麻美，趙　國，山下洋一，“アクセント結合規則を利用した統計的手法に基づく連続音声のアクセント型自動ラベリング,” 音響誌，vol.66, no.10, pp.487–496, 2010.

[9] 山本麻美，趙　國，山下洋一，“言語情報と F0 情報を利用したアクセント句境界の自動推定,” 信学技報，SP2010-109, 2011.

[10] 黒岩　龍，日本語音声合成のためのアクセント結合規則の改善とデータベースに基づく統計的アクセント処理，東京大学大学院修士論文，2007.

[11] 日本音響学会新聞記事読み上げ音声コーパス (JNAS), http://research.nii.ac.jp/src/JNAS.html

[12] 新聞記事読み上げ高齢者音声コーパス (S-JNAS), http://research.nii.ac.jp/src/S-JNAS.html

[13] 伝　康晴，小木曽智信，小椋秀樹，山田　篤，峯松信明，内元清貴，小磯花絵，“コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用,” 日本語科学，no.22, pp.101–122, 2007.

[14] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random felds: Probabilistic models for segmenting and labeling sequence data,” Proc. 18th International Conference on Machine Learning (ICML), pp.282–289, 2001.

[15] 印南圭祐，“CRF を用いた日本語アクセント結合処理における誤り解析とそれに基づく改良,” 東京大学大学院修士論文，2009.

[16] S. Sekine and H. Isahara, “IREX: IR and IE evaluation project in Japanese,” Proc. LREC 2000.

[17] 小林俊平，条件付確率場に基づく日本語アクセント型予測モデルの改良と日本語教育システムへの応用，東京大学大学院修士論文，2012.

[18] MeCab, http://code.google.com/p/mecab/

[19] CaboCha/南瓜，http://code.google.com/p/cabocha/

[20] WP2TXT, http://wp2txt.rubyforge.org/

[21] CRF++, https://code.google.com/p/crfpp/

[22] NAIST Japanese Dictionary, http://sourceforge.jp/projects/naist-jdic/

[23] hts_engine API, http://hts-engine.sourceforge.net/

[24] MMDAgent, http://www.mmdagent.jp/

[25] 長野　徹，立花隆輝，西村雅史，“コーパスベース日本語音声合成フロントエンド,” 信学論（D），vol.J93-D, no.10, pp.2096–2106, Oct. 2010.

[26] 中田陽介，N. Graham，森　信介，河原達也，“点予測による形態素解析,” 情処学研報，2010-NL198, pp.1–7, 2010.

[27] 森　信介，“点予測による自然言語処理,” 第 8 回 Tokyo-NLP，2011.

（平成 24 年 6 月 3 日受付，9 月 24 日再受付）

鈴木　雅之　（学生員）

2010 東京大学大学院工学系研究科修士課程修了．修士（工学）．現在，同大学院工学系研究科博士後期課程に在籍．音声認識，音声強調，音声合成に関する研究に従事．IEEE，ISCA，情報処理学会，日本音響学会各会員．

黒岩　　龍

2007 東京大学大学院情報理工学系研究科修士課程了．修士（情報理工学）．現在，（株）NTT データ所属．

印南　圭祐

2009 東京大学大学院新領域創成科学研究科修士課程了．修士（科学）．現在，富士通（株）所属．

小林　俊平

2012 東京大学大学院情報理工学系研究科修士課程了．修士（情報理工学）．現在，（株）野村総合研究所にて生命保険会社向けシステムの開発に従事．

清水　信哉

2012 東京大学大学院情報理工学系研究科修士課程了．修士（情報理工学）．自然言語処理，音声認識に関する研究を行う．現在，マッキンゼー・アンド・カンパニー・インク・ジャパンに在籍．

峯松　信明　（正員）

1995 東京大学大学院工学系研究科博士課程了．博士（工学）．現在，同大学院工学系研究科教授．2002～2003 在外研究員（KTH，スウェーデン）．科学から工学に至るまで，音声コミュニケーションに関する研究に従事．IEEE，ISCA，SLaTE，IPA，CALICO，音響学会，情報処理学会，人工知能学会，音声学会，音声言語医学会，外国語教育メディア学会各会員．

広瀬　啓吉　（正員：フェロー）

1972 東大・工・電気工学卒．1977 同大大学院博士課程了．工博．同年東京大学工学部電気工学科講師．1994 同電子工学科教授．1996 東京大学大学院工学系研究科電子情報工学専攻教授．1999 同新領域創成科学研究科教授．2004 年 10 月より同情報理工学系研究科教授．1987 米国 MIT 客員研究員．音声言語情報処理分野一般についての研究開発に従事，特に韻律に着目した研究．IEEE，米国音響学会，ISCA（Board メンバー），情報処理学会，日本音響学会，人工知能学会，言語処理学会，信号処理学会各会員．