Vol.2021-MUS-131 No.48
Vol.2021-SLP-137 No.48
2021/6/19

# 日本語アクセント結合推定における LSTMと単語埋め込み表現の利用

角南 陽友[1]　齋藤 大輔[1]　峯松 信明[1]

**概要**：近年，入力テキストから直接音声波形をネットワークを通して生成する End-to-End 型の音声合成が主流となっている．この場合，漢字の読みからアクセント変形などの韻律制御まで，全てを入力テキスト（漢字仮名混じり文）から直接的に推定することになるが，推定誤りは避けられない．その一方で，従来より行われているように，読みや韻律制御を事前にテキスト処理によって推定し，入力を「韻律シンボルが適宜埋め込まれた平仮名列」として合成器に入力する手法も検討されている [1]．本研究では，波形生成処理の前処理である，テキスト処理としての韻律推定（主としてアクセント結合推定）について報告する．筆者らの先行研究では，入力単語列の各種言語素性を用いて，CRF によりアクセント結合後のアクセント位置を推定していたが，単語の意味的な関係に起因するアクセント変形の対処（特に複合名詞）が不十分であった．本研究ではこれらに対処することを目的として，LSTM ニューラルネットワークと Word2Vec および BERT による単語埋め込み表現を利用し，推定精度の向上を実現したので，その結果を報告する．

## 1. はじめに

近年の音声合成システムの性能向上に伴い，様々な場面で音声合成が使用されるようになった．公共施設などでの音声案内をはじめとし，スマートスピーカーやエンターテインメイント目的でも利用されるようになったテキスト音声合成システムは，今後も技術の進歩に伴って需要が高まり，さらに自然な音声を合成することが求められるようになると予想される．

テキスト音声合成に必要な情報としては，テキストから G2P 変換を通して得られる音素列や，音高を表す韻律情報がある．韻律情報は，単語ごとに固有のアクセントと，句や文のレベルでの音高の大域的な変化を表すイントネーションとに分けられる．イントネーションは話者や発話状況に依存して変化するが，アクセントは同一方言であれば話者によらず同じであり，誤ったアクセントで読み上げると不自然な日本語になってしまう．特に，日本語では「雨」と「飴」など，音素列が同じであってもアクセントによって弁別される単語があり，また，prince としての「王子」と地名としての「王子」のように表記が同じであっても，アクセントで意味を弁別することもある．したがって，テキスト音声合成において自然な日本語の読み上げを実現するには，正しいアクセント情報を付与することが重要である．

[1] 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo

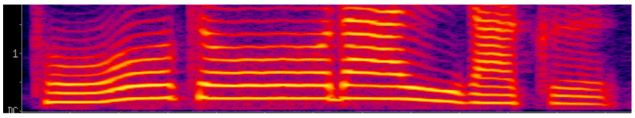


**図 1**　「東京大学」の狭帯域スペクトログラム．横縞がピッチのハーモニクスに対応する．「ダ」の直後で音高が下降している．
**Fig. 1**　Spectrogram of ”Tokyo Daigaku”.

### 1.1 日本語のアクセント

日本語におけるアクセントは，モーラ単位で音の高さが変化することによる高低アクセントである（**図 1**）．特に東京方言については，アクセントを定式化するための様々な議論がなされ，アクセントの定義自体も変化してきた [2]．各モーラの音高の高低によりアクセントを分類する考え方に始まり，昭和初期には，アクセントを各モーラの音高の高低ではなく音高の上昇・下降に着目して捉える考え方が登場した．その後，音高の上昇に関して，東京方言では全ての語で共通して冒頭で上昇するという特徴があることから，単語固有のアクセントとしては音高の下降する部分が注目されるようになった．

本研究では，単語中で音高が下降する直前のモーラをアクセント核とし，N 番目のモーラがアクセント核である単語を N 型であると分類する．また，音高の下降がない単語は 0 型であるとする．

単語が連接して文節や複合単語となる際，単独で発声す

ⓒ 2021 Information Processing Society of Japan

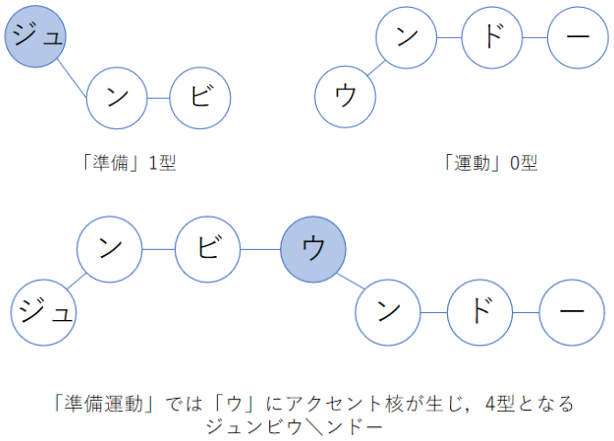


図 2　アクセント結合の例「準備運動」
**Fig. 2**　An example of accent sandhi in “Junbi Undo”.

る場合とは異なるアクセントに変化する現象がしばしば起こり，このようなアクセントの変化をアクセント結合と呼ぶ（図 2）．音高が上昇し下降するというアクセント的にひとまとまりになった単位をアクセント句と呼ぶ．

本研究の目的は，辞書に収録された単語ごとの単独発声時のアクセント型から，文中で変化したアクセント型を推定することである．そのために，アクセント句境界の推定とアクセント核位置の推定という 2 つのタスクを行う．

類似の研究として，ニューラルネットワークを用いてアクセントを推定し，NEologd[3] をベースとした大規模なアクセント辞書を作成する研究が行われている [4]．複合単語も含んだアクセント辞書を作成することにより，辞書に存在する単語については正しいアクセント情報を得ることができるが，辞書にない複合単語や，付属語との結合では別途推定する必要がある．また，アクセント辞書を作成する段階ではアクセント結合推定が必要になり，その部分の精度を向上させることができる点でも本研究の意義がある．

## 2.　関連研究

### 2.1　規則によるアクセント推定

単語が連結することによるアクセントの変化を統一的に扱う試みとして，匂坂らにより付属語連鎖および複合単語のアクセント結合を規則化が行われている [5]．

さらに特殊な場合についてのアクセント結合規則として，数詞のアクセント結合が挙げられる．宮崎は数詞と助数詞の組み合わせによるアクセント変化に着目し，数詞および助数詞をグループ化し，数詞に助数詞が結合した場合のアクセント変化を分類した [6]．

### 2.2　条件付き確率場を用いたアクセント結合推定

機械学習を利用したアクセント結合推定手法として，鈴木らは，入力テキストから様々な言語素性を抽出し，条件付き確率場（Conditional Random Field; CRF）を用いた統計的な学習によるアクセント推定を試みた [7]．

アクセント推定は，日本語の文の各形態素にアクセント型を付与するという系列ラベリング問題として定式化できる．CRF は，系列ラベリング問題を解くのに利用される識別モデルである．入力データ列 $\mathbf{x} = x_1, x_2, \ldots, x_N$，ラベル列 $\mathbf{y} = y_1, y_2, \ldots, y_N$ に対して，ラベル系列の事後確率を，下式でモデル化する．

$$P(y|\mathbf{x}) = \frac{\exp \sum_{f=1}^{F} w_f \phi_f(\mathbf{x}, y)}{\sum_{y' \in Y} \exp \sum_{f=1}^{F} w_f \phi_f(\mathbf{x}, y')} \tag{1}$$

ここで，$\phi_f(\mathbf{y}|\mathbf{x})$ は素性関数と呼ばれる関数であり，素性 $f$ の値が一致するとき 1，一致しないとき 0 をとる．$w_f$ は素性 $f$ に対応する重みであり，

$$\mathcal{L} = \sum_i \log\left(P(y_i|\mathbf{x})\right) \tag{2}$$

を最大化するように学習される．

### 2.3　アクセントの谷

前述（2.1）のアクセント結合規則は，アクセント結合の際にアクセント型がどのように変化するかについて論じたものであり，アクセント結合が起きるかどうかという点については規則化されていなかった．一方，定延は，複合語内で音高が下降し再び上昇する部分を「アクセントの谷」と定義し，その生起条件について検討している [8]．

まず，接頭辞付加による複合語では，後部要素の字数がアクセントの谷の生起に影響し，例えば接頭辞「全」について，「全公共団体」「全消費者」「全社員」など後部要素が長いものではアクセントの谷が生じるが，「全社」「全党」「全員」などではアクセントの谷は生じない（アクセント結合が起こる）．ただし，「全自動」のように語彙化の度合いが高いものは，後部要素が 2 字以上であっても例外的にアクセント結合が生じる．「全」と同様に後部要素が 2 字以上の場合にアクセントの谷が生じやすい接頭辞として，「各」「前」「対」などが挙げられている．しかし，後部要素の字数によるアクセントの谷の生起パターンは，接頭辞の種類によって異なり，「第」では後部要素が 1 字であってもアクセントの谷が生じやすく，一方で「非」は後部要素が 3 字以上の場合にアクセントの谷が生じやすい．

また，接頭辞以外の単語が複合した複合語中でアクセントの谷の生起が可能な構成要素間の関係について，以下の場合が挙げられている．ここでは，「選手宣誓」のように主語+動詞のパターンではアクセント結合が起こらない可能性がある，といった例が示されている．

(1) 主語 + 動詞　：選手宣誓，攻守逆転
(2) 主語 + 形容（動）詞 ：意識不明，意味深長
(3) 目的語 + 動詞　：自信喪失，内閣改造
(4) 同義　：拍手喝采，大義名分
(5) 同位　：公平中立，呆然自失

(6) 反意　：保守革新，適不適
(7) 人名（姓+名）　：須藤耕作，東郷敬治
(8) 組織名 + 役職名　：政府要人，自治会会長
(9) 氏名+地位・(役) 職名: 加藤教授，須藤助手
(10)順番を表す名詞+地位・役職名: 初代会長，次期大統領
(11)地域名+地域をさらに指定する名詞 ：近畿南部，本州最南端
(12)目標を表す名詞+動詞 ：当駅到着，本校着任
(13)場所を表す名詞+動詞 ：都内在住，合衆国滞在
(14)時間を表す名詞+動詞 ：近日上映，早期発見
(15)役職名+氏名 ：捕手斉藤
(16)場所や時間の名詞+順位や評価の名詞 ：本邦最初，今世紀最大
(17)名詞+独特・特有・専用 ：西洋人独特，東洋特有，社長専用
(18)場所を表す名詞+名産・周辺・所収 ：当地名産，皇居周辺，新年号所収
(19)形容されるものを表す名詞+形容する擬音語・擬態語 ：非難轟々，虎視眈々
(20)要+動作を表す名詞 ：要検討，要修理
(21)その他 ：感無量，観自在，準備万端

また高野らは，4 つ以上の形態素からなる長い複合名詞について，単語間の結合順序を表す木構造を手動で作成し特徴量とすることにより，後述の CRF によるアクセント結合推定の精度が向上したことを報告している [9]．例として「専門家検討会」は (専門+家)+(検討+会)，「日本弁護士連合会」は日本+(弁護士+(連合+会)) のような順で複合しており，木構造の上位の節点にアクセントの谷が生じる．このような構成要素間の統語的な関係がアクセント結合の有無を判断するにあたって有用であることを示している．しかし，木構造自体は自動推定ではなく手動で設定したものであり，統語的な関係を利用してアクセント結合の自動推定を行うには至っていなかった．

## 3. 提案手法

前節で示したように複合名詞においては，複合名詞内の名詞間の統語構造や，意味的関係，更には各単語の意味属性に起因する形で，複合名詞に（例外的な）「アクセントの谷」が生起する様子が説明されている．そこで本研究では，これらの例外処理に対応できるよう，深層学習を取り入れたアクセント結合推定を行った．モデルとしては時系列データに使われる LSTM ニューラルネットワークを用い，さらに，単語間の意味的関係を捉えるため，Word2Vec および BERT を用いた単語埋め込み表現を特徴量に加えた．

### 3.1 LSTM ニューラルネットワークの利用

時系列データを処理する深層学習モデルとして，再帰型ニューラルネットワーク (Recurrent Newral Network; RNN) が用いられる．RNN は時系列データに対して用いられるネットワークであり，中間層の出力を再帰的に入力することで，系列中の離れたデータの影響も考慮した出力を得ることができるとされている．本研究では，RNN を改良した LSTM ニューラルネットワークを使用し，実装には PyTorch を利用した．[*1]

### 3.2 Word2Vec による単語埋め込み表現

機械学習において自然言語を扱うには，単語をベクトルとして表現する必要がある．最も単純な手段は，各単語に対する one-hot ベクトルを利用する方法であるが，語彙数が増えると膨大な次元数が必要になってしまう．また，one-hot ベクトルでは単語を文字列として数値に対応させているだけで，単語の意味を表現することはできていない．

Mikolov ら [10] によって考案された Word2Vec は，ニューラルネットワークに基づき，単語をベクトル空間上で表現する手法である．これによって単語はベクトルとして表現され，単語の分散表現や埋め込み表現 (word embedding) と呼ばれる．本研究では日本語版 Wikipedia の全文を学習データとして使用した日本語 Wikipedia エンティティベクトル[*2] を Word2Vec として利用した．

### 3.3 BERT による単語埋め込み表現

最近の自然言語処理で注目されている技術として，2018 年に Google により開発された BERT(Bidirectional Encoder Representations from Transformers) がある [11]．BERT は文意の把握に優れているとされ，自然言語処理の様々なタスクで SoTA を達成している．

BERT の基礎となっているのは，Transformer と呼ばれるエンコーダ・デコーダモデルである [12]．Transformer では，離れた位置にあるデータの関連性を示す Self-Attention と呼ばれる機構を自然言語処理に応用し，離れた位置にある着目すべき単語を考慮できるとされている．BERT の構造は Transformer のエンコーダを重ねたものであり，最終層の出力として単語ごとのベクトル表現が得られる．

BERT による単語埋め込み表現では，出現位置や周辺の単語との関係によって同じ単語でも異なるベクトル表現が得られる．**図 3** は，「運動」という語を含む複合単語について BERT による単語埋め込み表現を取得し，三次元にプロットしたものである．「運動○○」「○○運動」と「運動」以外の単語が別々の位置にプロットされていることがわかる．「運動」という単語のアクセント型は「運動○○」では 0 型または 3 型，「○○運動」では 1 型に概ね対応しており，単語の出現環境によって異なるベクトルが得られ

*1 PyTorch
https://pytorch.org/
*2 日本語 Wikipedia エンティティベクトル
http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

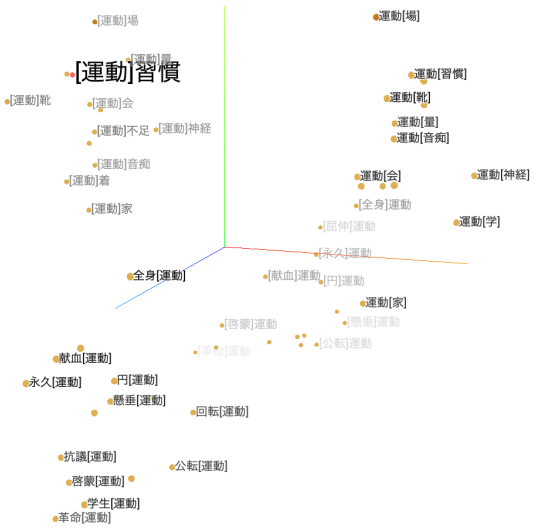


図 3　BERT による単語埋め込み表現
**Fig. 3**　Word-embedding with BERT

る BERT を特徴量として利用することでアクセント推定性能が向上することが期待される．本研究では，BERT の事前学習モデルとして NICT BERT 日本語 Pre-trained モデル[*3] を利用した．

## 4.　評価実験

アクセント句境界およびアクセント核位置の推定実験を行った．学習データ・評価データには鈴木ら [7] と同様に日本音響学会新聞記事読み上げ音声コーパス（JNAS）を利用した．全 6334 文に対し，単独のラベラによりアクセント句境界・アクセント核位置がラベリングされている．ラベルデータに含まれる文に対し形態素解析器 MeCab[*4]により形態素解析を行い，形態素の境界・読みがラベルデータと一致しなかったものは除外した．形態素解析用の辞書としては現代話し言葉 UniDic[*5]を用いた．使用できるデータは 4670 文となり，そのうち 999 文を評価データとし，残りの 3671 文を訓練データとした．

アクセント句境界推定・アクセント核位置推定のそれぞれについて，まず既存の言語特徴のみを利用し，CRF を用いた場合と LSTM を用いた場合とで学習・推定を行った．詳細は [7] に譲るが，既存の言語特徴としては品詞，活用型，活用形，モーラ数，単独発声時のアクセント型，アクセント修飾型，アクセント結合型，数詞・助数詞のタイプなどを形態素解析結果をもとに取得し利用している．次に LSTM モデルにおいて，特徴量に Word2Vec または BERT による単語埋め込み表現を追加し，学習・推定を行った．また，LSTM に Word2Vec を加えた場合が最も性能が良かったため，追加実験として，LSTM を二層重ねたモデルも検討した．また，本研究では複合名詞のアクセントに着目したため，名詞連続部分のみについての評価も行った．

*3　NICT BERT 日本語 Pre-trained モデル
https://alaginrc.nict.go.jp/nict-bert/index.html
*4　MeCab https://taku910.github.io/mecab/
*5　UniDic
https://unidic.ninjal.ac.jp/

表 1　アクセント句境界推定精度の比較 (全形態素)

| モデル | 単語埋め込み | Recall | Precision | F 値 |
|---|---|---|---|---|
| CRF | - | 90.52 | **97.29** | 93.79 |
| LSTM | - | 95.53 | 94.32 | 94.92 |
| LSTM | Word2Vec | **96.23** | 93.97 | **95.09** |
| LSTM | BERT | 95.78 | 92.65 | 94.19 |
| LSTM(2 層) | Word2Vec | 96.03 | 94.44 | 95.23 |

表 2　アクセント句境界推定精度の比較 (名詞連続部分)

| モデル | 単語埋め込み | Recall | Precision | F 値 |
|---|---|---|---|---|
| CRF | - | 66.56 | **87.04** | 75.43 |
| LSTM | - | 73.99 | 74.92 | 74.45 |
| LSTM | Word2Vec | 80.96 | 75.80 | **78.29** |
| LSTM | BERT | **84.67** | 71.60 | 77.59 |
| LSTM(2 層) | Word2Vec | 83.75 | 76.96 | 80.21 |

### 4.1　アクセント句境界推定

アクセント句境界推定の評価尺度として，Recall，Precision，F 値を用いた．Recall（再現率）は，アクセント句境界が実際に存在するもののうち，正しく推定されたものの割合である．Precision（適合率）は，推定されたアクセント句境界のうち，実際に存在するものの割合である．全ての形態素にアクセント句境界が存在するなど極端な推定でも Recall が高くなってしまい，逆に極端に少なく推定しても Precision が高くなってしまう場合があるため，総合的な性能の評価としては Recall と Precision の調和平均である F 値が用いられる．結果を**表 1** および**表 2** に示す．

### 4.2　アクセント核位置推定

アクセント核位置推定では，単独発声時のアクセントからどのようにアクセント型が変化するかという相対変化アクセントラベルを分類することにより学習・推定を行った．相対変化アクセントラベルの種類は次の通りである．

“non” 単独発声時にはアクセント核があるが，文中では無核．
“same” 単独発声時にアクセント核があり，アクセント結合後も同じ位置にアクセント核がある．
“same-1” 単独発声時にアクセント核があり，アクセント結合で 1 つ前のモーラに移動する．
“samenon” 単独発声時に無核で，アクセント結合後も無核．
“mora” 末尾のモーラがアクセント核になる．
“atama” 先頭のモーラがアクセント核になる．
“mora-1” 末尾の 1 つ前のモーラがアクセント核になる．
いずれにも当てはまらない場合，アクセント結合前後での

表 3　アクセント核位置推定精度の比較 (全形態素)

| モデル | 単語埋め込み | 正解率 |
|---|---|---|
| CRF | - | 94.92 |
| LSTM | - | 94.68 |
| LSTM | Word2Vec | **95.05** |
| LSTM | BERT | 92.88 |
| LSTM(2 層) | Word2Vec | 95.24 |

表 4　アクセント核位置推定精度の比較 (名詞連続部分)

| モデル | 単語埋め込み | 正解率 |
|---|---|---|
| CRF | - | 89.31 |
| LSTM | - | 87.19 |
| LSTM | Word2Vec | **89.82** |
| LSTM | BERT | 86.34 |
| LSTM(2 層) | Word2Vec | 90.48 |

核移動幅をラベルとした．また，複数の条件に合致する場合は，より上に記述したラベルを採用した．結果を**表 3** および**表 4** に示す．アクセント型の正解率で評価した．

### 4.3　推定結果の分析

本研究で使用したデータにおける推定結果を分析し，改善点について，その傾向を確認した．また，最も性能の高かった単語埋め込み表現として Word2Vec を使用した 2 層の LSTM モデルについて，実際の推定結果を分析した．

まず，CRF を使用した既存手法では誤っていたが LSTM で正解した例を表 5 に示す．アクセント句境界があると判定された部分をスラッシュ（/）で表し，推定されたアクセント核が何番目のモーラであるかを各単語の後の括弧内に示した．表 5 において，a〜g は 2.3 節において，アクセント結合が起こらない，「アクセントの谷」が出現するパターンに該当する．例えば，a「意識過剰」は「意識 (が) 過剰 (だ)」という主語+形容 (動) 詞という関係に，d「前期後期」は反意の語を並べた関係になっており，いずれもアクセント結合が起こらない．既存手法ではこれらの場合にアクセント結合が起こると推定し誤っていたが，提案手法ではアクセント句境界を予測し，正しいアクセントが得られている．これらの例では，複合単語の構成要素間の意味的な関係を捉えることができたと考えられる．

h「当該国刑事裁判法廷」, i「国内 B 級ライセンス取得」は，いずれも長い複合名詞である．これらの例では，CRF においてはアクセント句境界が誤っており，単語の結合順が正しく推定されていなかった．しかし提案手法では，アクセント句境界が正しく推定され，正しいアクセントが得られている．これは，長い名詞連続における単語間の統語的な関係を捉えることができたと考えられる．

また，本研究で直接検討してはいなかったものではあるが，j「十三万七千円」のような長い数詞に関しても改善が見られた．これについては，LSTM により離れた単語の影響を考慮できるようになった結果であると考えられる．

表 5　アクセント結合推定結果が改善された例

| | 上段：CRF での推定結果（誤り）<br>下段：LSTM での推定結果（正解） |
|---|---|
| a | /意識 (0) 過剰 (0) に (0) |
| | /意識 (1)/過剰 (0) に (0) |
| b | /子宮 (0) 全体 (1) が (0) |
| | /子宮 (0)/全体 (0) が (0) |
| c | /野党 (0) 多数 (2) の (0) |
| | /野党 (1)/多数 (2) の (0) |
| d | /前期 (0) 後期 (1) |
| | /前期 (1)/後期 (1) |
| e | /八十 (0) 年 (0) 代 (0) 後半 (1) |
| | /八十 (0) 年 (1) 代 (0)/後半 (0) |
| f | /推進 (0) 費 (0) 盛り込み (1) を (0) |
| | /推進 (3) 費 (0)/盛り込み (0) を (0) |
| g | /特別 (0) 査察 (0) 受け入れ (1) に (0) |
| | /特別 (0) 査察 (1)/受け入れ (0) に (0) |
| h | /当該 (0) 国 (0) 刑事 (1)/裁判 (0) 法廷 (1) が (0) |
| | /当該 (3) 国 (0)/刑事 (0) 裁判 (0) 法廷 (1) が (0) |
| i | /国内 (2)/B (0) 級 (0)/ライセンス (0) 取得 (1) |
| | /国内 (2)/B (0) 級 (0)/ライセンス (1)/取得 (0) |
| j | /十 (0) 三 (1) 万 (0)/七千 (0) 円 (0) の (0) |
| | /十 (0) 三 (0) 万 (1)/七千 (0) 円 (0) の (0) |

次に，LSTM で正しく推定できなかった例を表 6 に示す．LSTM を用いた場合，CRF を用いた手法に比べて Precision が低いという結果になっており，実際に LSTM での誤りの多くが，アクセント句境界が存在しない部分に対しアクセント句境界があると推定する挿入誤りであった．a「日米（の）関係」，b「連続（した）放火」のような修飾語+被修飾語の関係は，2.3 節のアクセントの谷の生起条件に該当せず，アクセント結合が起こる例である．この誤りを列挙したところ，複合名詞の構成要素のうち最後の単語のみが後続の付属語と同じアクセント句を構成することによる誤りが多く見られた．その傾向は e「回線接続問題が」，f「先進国経済の」のような三単語の名詞連続で明確に現れていた．これについて，統語的には最後の単語が複合単語の前部要素よりも後続の付属語との繋がりが強い，すなわち，その単語が複合名詞を構成する要素であることを正しく捉えられていないと考えられる．

## 5.　まとめと今後の課題

これまで使用されてきた CRF を用いた手法に代わり LSTM ニューラルネットワークを利用し，特徴量として Word2Vec による単語埋め込み表現を導入することで，アクセント句境界推定・アクセント核位置推定のいずれにおいても性能が向上した．特にアクセント句境界推定において，単語埋め込み表現を用いた場合に Recall（再現率）の値が向上していた．複合名詞においてアクセント句境界の再現率はアクセント結合の起こらない「アクセントの谷」を正しく推定した割合であるといえる．実際の推定結果か

表 6 LSTM によるアクセント結合推定結果が誤った例

| | 上段：LSTM での推定結果（誤り）<br>下段：正解ラベル |
|---|---|
| a | /日米 (1)/関係 (0) など (1) を (0)<br>/日米 (0) 関係 (1) など (0) を (0) |
| b | /連続 (0)/放火 (0) の (0)<br>/連続 (0) 放火 (1) の (0) |
| c | /従来 (1)/モデル (0) より (0)<br>/従来 (0) モデル (1) より (0) |
| d | /カバー (1)/範囲 (1) を (0)<br>/カバー (0) 範囲 (1) を (0) |
| e | /回線 (0) 接続 (0)/問題 (0) が (0)<br>/回線 (0) 接続 (0) 問題 (1) が (0) |
| f | /先進 (3) 国 (0)/経済 (1) の (0)<br>/先進 (0) 国 (0) 経済 (1) の (0) |

らも，アクセント結合が起こらない位置を正しく認識していることがうかがえる．また，長い複合名詞の句境界や長い数詞句アクセントについても改善がみられ，LSTM を利用したことで離れた単語の影響を利用できた結果であると考えられる．また，LSTM のレイヤー数を 1 層から 2 層に増やした場合にはさらに性能が向上したことから，モデルのパラメータも性能に影響を与えると考えられる．

一方，単語埋め込み表現として BERT を用いた場合，周辺の語の影響をより考慮して推定できると考え精度の向上を期待したが，全体として Word2Vec に比べて低い性能を示し，特にアクセント核位置推定については，CRF を用いた既存手法や単語埋め込み表現を用いない場合よりも正解率が下がってしまう結果となった．これについては，アクセント結合が生じるか生じないかには離れた単語も含めた統語的関係も影響するが，アクセント結合後のアクセントの変化に関しては直接隣り合った単語の影響が大きく，離れた単語を考慮した単語ベクトルによってかえって正しくない推定が行われてしまったのではないかと考える．隣接する単語の影響が大きいアクセント核位置推定では畳み込みニューラルネットワーク（CNN）を利用するなど，タスクの特徴によって異なる構造のモデルを利用することも今後検討すべき事項である．

なお，本研究で対象としたアクセント結合に関しては BERT では性能向上がみられなかったが，「ネット（網/インターネット）」「ライブ (生放送/生演奏)」「メモリー (思い出/記憶装置)」「王子 (Prince/地名)」のように，文脈によりアクセントが変化する単語のアクセント分類では，周辺の語の文脈を考慮する BERT の性質をより発揮できるのではないかと考える．

提案手法での誤りの多くは，複合名詞の範囲を誤って捉えていたものであった．本研究では全ての単語について同じモデルで推定を行ってきたが，複合名詞部分のみを取り出して推定を行うなど，単語の出現環境に応じて最適化されたモデルを構築することも検討すべきである．

単語埋め込み表現以外の特徴量や学習データは先行研究で使用されていたものを引き継いで使用したが，改めて使用する特徴量の検討を行うことや，特に例外的な場合について学習データを増やすことは改善に繋がると考える．

また，推定結果において誤りと判定されたものの中には，辞書的な読みから規則的に推定されるアクセントとは異なるものの，許容範囲であると考えられるものもあった．「米国史」は「ベイコク＼シ」で 4 型であるが，実際には「ベイコッシ」のように母音が無声化して発音される場合があり，3 型と推定されても許容範囲であると考えられる．許容できない誤りと許容範囲の誤りを区別して評価することも必要であろう．

## 参考文献

[1] 栗原清，清山信正，熊野正，今井篤：読み仮名と韻律記号を入力とする日本語 end-to-end 音声合成の音質評価，信学技報，Vol. 118, No. 354 (2018).
[2] 杉原満：音声表現から見る共通語の韻律理論–『NHK アクセント辞典』改訂に向けて，放送研究と調査，Vol. 61, No. 4, pp. 76–90 (2011).
[3] 佐藤敏紀，橋本泰一，奥村学：単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討，言語処理学会第 23 回年次大会，Vol. NLP2017, pp. 875–878 (2017).
[4] Tachibana, H. and Katayama, Y.: Accent Estimation of Japanese Words from Their Surfaces and Romanizations for Building Large Vocabulary Accent Dictionaries, *ICASSP 2020*, pp. 8059–8063 (2020).
[5] 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 D， Vol. 66, No. 7, pp. 849–856 (1983).
[6] 宮崎正弘：日本文音声変換のための数詞読み規則，情報処理学会論文誌，Vol. 25, No. 6, pp. 1035–1043 (1984).
[7] 鈴木雅之，黒岩龍，印南圭祐，小林俊平，清水信哉，峯松信明，広瀬啓吉：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，電子情報通信学会論文誌. D, 情報・システム， Vol. J96-D, No. 3, pp. 644–654 (2013).
[8] 定延利之：現代日本語東京方言における合成的字音語のアクセントと字数について，*LEBULAE*, Vol. 15, No. 12, pp. 140–157 (1991).
[9] 高野克弥，清水信哉，峯松信明，広瀬啓吉：複合名詞内アクセント句境界を用いたアクセント結合予測の高精度化に関する実験的検討，日本音響学会研究発表会講演論文集，Vol. 2011-Mar, No. 1-Q-12(d), pp. 363–364 (2011).
[10] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *International Conference on Learning Representations*, (online), available from ⟨http://arxiv.org/abs/1301.3781⟩ (2013).
[11] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *NAACL 2019*, Vol. 1, pp. 4171–4186 (2018).
[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention Is All You Need, *NIPS 2017*, (online), available from ⟨https://arxiv.org/pdf/1706.03762.pdf⟩ (2017).

ⓒ 2021 Information Processing Society of Japan