特定話者による大規模アクセントラベリングとそのデータベース化*

◎黒岩 龍（東大・情報理工）, 峯松信明（東大・新領域）,
△伝 康晴（千葉大・文）, 広瀬啓吉 (東大・情報理工)

1 はじめに

日本語テキスト音声合成ソフトウェアに漢字仮名混じり文のテキストが入力された場合、一般的には、まず付属する形態素辞書を参照してテキスト解析（形態素解析等）を行なうことによって形態素境界や各形態素の読み（モーラ列もしくは音素列）・基本アクセント型などの情報を得、それを基に音韻処理（無声化などを含む処理）と韻律処理（アクセント・イントネーション・継続長・パワーなどの処理）がそれぞれ行なわれ、最終的な波形出力となる。辞書に登録されているアクセントの情報は各形態素を単独発声したときのものであり、後の韻律処理において、文としてのアクセントを得るための的確な処理（アクセント結合）が必要となる。アクセント結合処理を規則によって表現し、それを基に日本語テキスト音声合成システムとして実装する研究がなされている [1, 2]。規則に基づいてすべての事象を網羅することには限界があるものの、統計的学習・推定を行なうために必要となる自由に利用できる多量の正解ラベル付きデータベースは存在しておらず、難しい状況にある。また、同じ理由で、規則による処理がどの程度妥当であるかも判定することができない問題もある。

以上のような問題を解決するため、大規模なアクセントラベリングを計画し、現時点において計画のおよそ四分の一が完成している。

2 特定話者による大規模アクセントラベリングデータベース

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、文はいくつかのまとまりに分かれ、それぞれの内部では音の高さが連続的に変化する。まとまりが始まる箇所ではピッチ（音の高さ）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。また、まとまりの内部には、語彙に依存した比較的急激な下降箇所がおおむね高々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できるわけではなく、複数のアクセント句が影響を及ぼし合い、また、融合する現象が見られる。これについて、日本語話し言葉コーパス [3] においては、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 つのアクセント核しか存在し得ない」との考えに基づき、複数のアクセント句として扱っている。

また、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かかわらず実際のピッチ変化に明確に現れていない場合が存在する [3]。その上、発声者等の知覚も必ずしも明確ではない。

このように、アクセント句とアクセント核はいずれも明確に認定できるものでないが、何らかの定義が必要となる。今回のアクセントラベリングにおいては以下のように定義し、これらをラベリング対象とした。

アクセント句 句頭上昇が見られる箇所をアクセント句の境界とする。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、通常はアクセント句の境界が生じる。

アクセント核 音の高さが下がる箇所の直前のモーラ。アクセント句内の出現数は制限しない。意識的に当該箇所で急激に音の高さを下げそれ以外では同じ音の高さで平坦に（イントネーションを除去して）読むようにした場合に、違和感が生じない位置をアクセント核と認める。

実際のラベリングでは、文中におけるアクセント句境界位置とアクセント核位置のラベリングを、1 人のみのラベラに行なわせるものとした。また、文中でのアクセントと個々の形態素を単独で発声する場合のアクセントを比較することによる研究で用いることを想定し、文中に出現するすべての自立語に対して単独発声時のアクセントについてのラベリングを行なわせた。形態素単独発声時のアクセントはアクセント辞典等に記されているが、アクセント感覚の個人差を含まないものとするため、文に対するラベリングを行なったラベラ自身に、改めて単独発声時についてのラベリングをさせた次第である。

すなわち、ラベリング対象となるのは、

- 文発声時についての “アクセント句境界位置” と “アクセント核位置”
- 文中に現れる自立語の単独発声時についての “アクセント核位置”

であり、これらをすべて 1 人のラベラが行なう。自立語でない形態素については、単独発声を想定することは難しく、また、不合理であるため行なっていない。

2.2 使用した文セット

日本音響学会 新聞記事読み上げ音声コーパス（JNAS）[4] で使用されている文（毎日新聞記事から選ばれた 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ることから、すべての文におよ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による。

ただし、不自然な読みや誤った読みも散見されるため、事前にすべての文の読みを確認し、適切でない箇所に対しては修正を施した。

2.3 形態素解析

文中に現れる自立語に関するラベリングをおこなうため、また、ラベリング結果を利用する際の利便性のため、すべての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[5] に基づくものとした。

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作業で修正した。従って、すべて正確に形態素解析が行われているわけではないが、読みについてはラベリングに用いたものと完全に同一になっている。

2.4 ラベリング従事者の選定

ラベリング作業に先立ち、ラベラとラベル検査担当者を選定した。

ラベリング作業をおこなうのは、先に述べた目的に合わせ、1 人のみとした。また、多量の文に対しラ

*A Large-scale Accent Labeling Experiment with a Single Labeler, by R. Kuroiwa (Univ. of Tokyo), N. Minematsu (Univ. of Tokyo), Y. Den (Chiba Univ.), K. Hirose (Univ. of Tokyo)

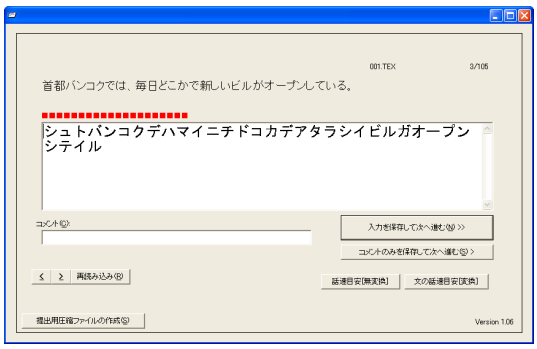

Fig. 1　ラベリング作業用プログラム

ベリング作業を行なうため誤ったラベルを付与してしまうことは免れ得ないため、ラベリングの済んだすべて文のラベルに目を通し、“誤りが含まれる可能性がある文” を選び出す作業の従事者（1 人, 今後増員の可能性あり）を、ラベラとは別に用意した。これら 2 人はいずれも東京生まれ・東京育ちの大学生であり、合唱の団体に所属するため比較的鋭い音感を持っていると推定される。ほぼ同様の学生 6 人の中から、アクセントに関する簡単な教育をした上で試験によって選抜したものである。

2.5　実際のラベリング作業

ラベリング作業は、音声を経ず、テキストに直接ラベルを付与させる形で行なった。音声を経ると極めて多くの作業時間を要するためであり、また、ラベラにとっての知覚的なアクセント情報を重視するためでもある。作業用に、Fig. 1 に示したプログラムを作成し、単純な操作でラベリング作業ができるようにした。

自立語のラベリングでは、完全に単独での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。

ラベル検査担当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、必要なものは訂正させた。

3　進捗状況と分析

3.1　進捗状況

現在、4,166 文（JNAS 新聞記事文の先頭 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、用意したすべての文に対してのラベリングを目指して進めていく。

3.2　分析

現時点で完成している 4,166 文での、アクセント句を構成する形態素の数が Table 1、アクセント句を構成する品詞列の上位 10 種類を表にしたものが Table 2 である。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。

Table 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%

Table 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%

4　おわりに

本論文では、アクセント研究のためのデータベース作成について解説し、現時点での進捗状況について報告した。以後、データの整理およびテキストの著作権等に関しての問題点を解決した上で、公開をする予定である。興味をお持ちの方にはその旨を一報されたい。

なお、本データベースを用いた統計的アクセント処理の研究については、[6] を参照していただきたい。

参考文献

[1] 匂坂, 佐藤:“日本語単語連鎖のアクセント規則”, 電子通信学会論文誌 J66-D.7, pp.847-856, 1983.

[2] 喜多 他: “日本語テキスト音声合成を目的としたアクセント結合規則の構築と改良”, 電子情報通信学会 信学技報, SP2002-26, pp.13-18, 2002.

[3] “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.

[4] 板橋 他: “日本音響学会新聞記事読み上げ音声コーパスの構築”, 日本音響学会 1997 年秋季講演論文集 I, pp.187-188, 1997.

[5] 伝 他: “話し言葉研究に適した電子化辞書の設計”, 第 2 回「話し言葉の科学と工学」ワークショップ講演予稿集, pp.36-46, 2002.

[6] 黒岩 他: “日本語音声合成を目的としたアクセント処理のための規則と統計的学習”, 日本音響学会 2007 年春季講演論文集 1-Q-19, 2007.