

日本語話し言葉コーパスの音声ラベリング

菊池 英明^{†,††} 前川喜久雄[†] 五十嵐陽介^{†,‡} 米山聖子^{†,‡‡} 藤本雅子[†]

[†] 国立国語研究所 ^{††} 早稲田大学人間科学部 [‡] 東京外国語大学大学院 ^{‡‡} 大東文化大学

1 はじめに

「日本語話し言葉コーパス (Corpus of Spontaneous Japanese, 以下 CSJ と呼ぶ)」には、音声言語の多様性分析に焦点を当てた研究における利用を視野に入れて、約 50 万語相当のデータ (時間にして約 45 時間) に対して人手により付加された分節音、韻律などの音声ラベル情報が格納される¹⁾。

これまでも音声データに対する音声ラベル情報付加は多数行われコーパスとして公開もされているがその多くは読み上げ音声を対象としたものである。CSJ に格納される自発音声に人手で音声ラベル情報を付加するためには、読み上げ音声に対する基準では扱えない問題を解決して自発音声に対する基準を準備することが必要となる。また、適用する基準の一貫性や安定性を高く保つことも要求される。本稿では、CSJ に対する分節音と韻律のラベル情報付加基準と、基準の妥当性を予備的に検証した結果を報告する。なお、以下ではラベル情報付加をラベリング、ラベリング作業者をラベラーと呼ぶ。

2 分節音ラベリング

2.1 ラベリング基準と方法

分節音のラベリングは、分節音ラベルを介して音声信号と言語記号とを時間軸上で対応づける作業である。ラベルとしては、音素相当の単位を基本としているが、音声学の利便性を考慮して一部に音素よりも音声に近い水準 (sub-phonemic level) のラベルを用意している。表 1 にラベリングに用いる記号を示す。表における口蓋化子音、<cl> (破裂音・破擦音の閉鎖区間) などが sub-phonemic level のラベルとして用意したものである。なお、“G”、“@” はそれぞれ摩擦音化、鼻音化した /g/ を表わすために用意したラベルであるが、ラベラーによる揺れが大きい CSJ の公開データには含まない可能性がある。その他、“H” は長母音、“F” は両唇摩擦音を表わす。CSJ のデータに対する分節音ラベル付加の例を図 1 の中に示す。

ラベリングの対象とする約 50 万語のデータは約 45 時間もの音声に相当する。ラベリングを効率よく実施するために、連続音声認識技術を応用した自動ラベリングを導入する。あらかじめ、さまざまな条件で CSJ の一部のデータに対する自動ラベリングを実施したところ、人手により付加したラベル位置との誤差の平均は最小で 14.3[ms] であった²⁾。この程度の精度でも人手によるラベリングの初期値として用いれば、初期値がない場合に比べて約 2 倍程度の効率向上が認められることが予備実験によりわかった。したがって、分節音ラベリングに際しては、自動ラベリングによって作成された初期値を人手により修正する形で作業を進める。なお、ラベラーには母音の無声化や子音の弱化、融合などによって境界を定めにくい場合に複数のラベルを一つに複合化して付与する処置を許している。ラベリングの結果に対して、基準の一貫性、安定性を高くするために、二重のチェックを行う。

2.2 ラベリングの精度

CSJ の一部のデータに対して、二人の熟練したラベラーが行ったラベリング結果を比較したところ、同一ラベル位置の誤差の絶対値は、講演によって多少のばらつきがあるものの全体で平均 8[ms] 程度であった²⁾。一般的に、読み上げ音声データに対する音素単位のラベリングにおいては誤差の平均を 5[ms] 程度に抑えるのを目標とすることが多いようである。話し言葉音声を中心とした CSJ のデータに対する分節音ラベリングでは、最終的な誤差を上記の 8[ms] 程度に抑えることを目標としてラベラーの訓練を行なった。

表 1: CSJ の分節音ラベル

カテゴリー	記号
母音	a,i,u,e,o,aH,iH,uH,eH,oH
無声化母音	A,I,U,E,O
子音	k,g,G,@,s,z,t,c,d,n,h,F,b,p,m,r,w
口蓋化子音	kj,gj,Gj,@j,sj,zj,cj,nj,hj
拗音系列	ky,gy,Gy,@y,sy,zy,cy,dy,ny,hy,by,py,my,ry
特殊記号	下記に別掲
特殊記号	定義
<cl>	破裂音・破擦音の閉鎖区間
<pz>	発話中の無音区間
<uv>	声帯振動終了後に継続している母音フォルマント
<sv>	母音終了後に継続している声帯振動
<v>	Vocal fry
<fv>	母音不確定音
<?>	子音不確定音
<N>	雑音
	呼吸音

表 2: X-JToBI の Tone ラベル

アクセントおよび句頭音調:	
A	アクセント核 (従来の H*+L)
H-	アクセント句頭の上昇音調
アクセント句の境界音調:	
L%	アクセント句終端で下降調の場合
%L	アクセント句始端
イントネーション句 (IP) の境界音調:	
H%	IP 終端で上昇調の場合
HL%	IP 終端で上昇-下降調の場合
LH%	IP 終端で下降-上昇調の場合
pH	上昇-下降調音調における F_0 ピーク
pL	下降-上昇調音調における F_0 上昇開始
<,>	音調がフラットになる区間
補助記号:	
x	ラベル位置での F_0 が信頼できない
!	ラベル位置が不明確
?	ラベルのカテゴリーが不明確

3 韻律ラベリング

3.1 ラベリングスキーム X-JToBI の提案

音声データの韻律情報を記述・表現するモデルはこれまでに数多く提案されている³⁾。日本語音声の韻律情報のラベリング手法としては ToBI (Tones and Break Indices) モデルに基づく J-ToBI と呼ばれるスキームが提案されており⁴⁾、読み上げ音声のラベリングスキームとして広く利用されている。

CSJ の自発音声の韻律ラベリングを実施するために、まず J-ToBI について、話し言葉音声に適用した結果を分析し、自発性の高い音声のラベリングにおいて種々の問題が生じることを明らかにした⁵⁾。そのうえで、その際の考察に基づいて J-ToBI を拡張し、新たな韻律ラベリングスキーム X-JToBI (eXtended J-ToBI) を提案した⁶⁾。X-JToBI の tone ラベルと BI ラベルについて、それぞれ表 2 と表 3 に示す。

また、CSJ のデータに対する X-JToBI のラベリングを行なった例を図 1 に示す。

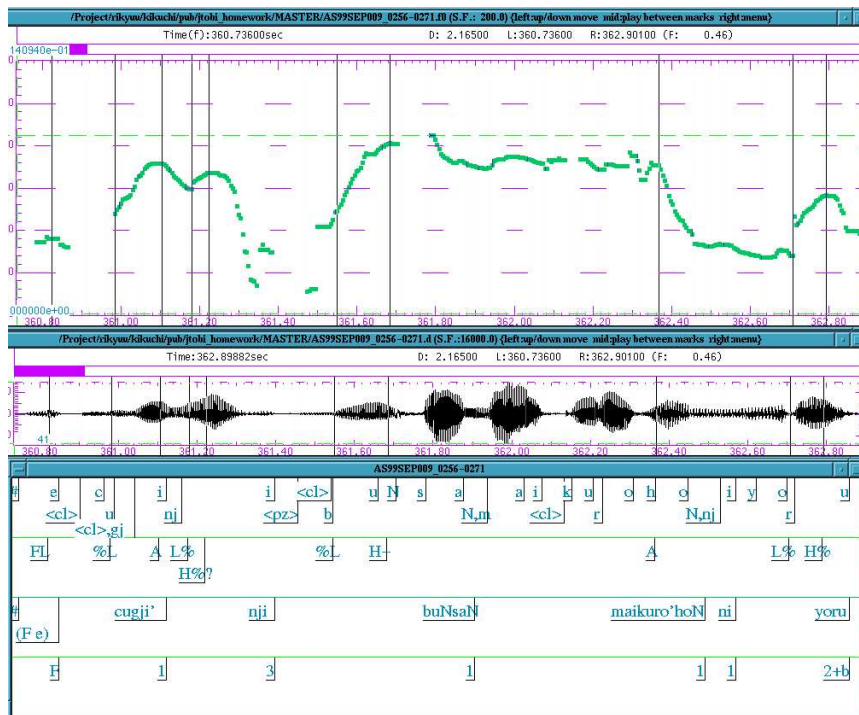


図 1: CSJ データに対する分節音および韻律ラベリングの例
 (ラベル層は上から、分節音ラベル、X-JToBI の tone 層、BI 層、Word 層。)
 (「え次に分散マイクロホンによる」)

表 3: X-JToBI の BI ラベル

(“OP” は Optional、“NR” は Not recommended を表す。)

BI ラベル	用法	備考
0	従来どおり	OP
1	従来どおり	
1+	1 と 2 の中間	NR
1+w	複合語内部の韻律境界	OP
1+p	アクセント句内部のポーズ	
2	従来どおり	
2+	2 と 3 の中間	NR
2+p	句末に BPM 無ポーズが後続ピッチレンジリセット無	
2+b	アクセント句末に BPM ポーズ後続せずピッチレンジリセット無	
2+bp	アクセント句末に BPM ポーズ後続ピッチレンジリセット無	
3	従来どおり	
3+	従来の”final”	OP
D	語断片ピッチリセット無	
D+	語断片ピッチリセット有	OP
P	語内部のポーズピッチリセット無	
P+	語内部のポーズピッチリセット有	OP
PB	境界音調の連鎖	
<F	韻律的フィラーの始端	
F	韻律的フィラーの終端	

3.2 ラベリングスキームの有効性

X-JToBI によって CSJ のデータのラベリングを行ない、ラベル出現頻度およびラベリング精度を調べた。その結果、X-JToBI で新規に導入したラベルが自発音声のラベリングに活用されることを確認した。また、その効果とラベリング基準の明確化によりラベルの正確性および再現可能性が向上することを確認した。詳しくは⁷⁾を参照されたい。

4 おわりに

本稿で示した基準にしたがって、現在までに CSJ の約 50 万語相当のデータに対して分節音および韻律ラベリングをほぼ完了した。来春のデータ公開に際してはラベリング用マニュアルも公開する予定である。なお、本稿で示したラベルの一部は公開に際して変更・廃止される可能性があるため注意されたい。

謝辞

分節音・韻律ラベリングの作業に従事していただいた方々に感謝申し上げます。

参考文献

- 1) 前川喜久雄, 『日本語話し言葉コーパス』の設計と実装, 平成 15 年度国立国語研究所公開研究発表会予稿集 (2003).
- 2) 菊池英明, 前川喜久雄, ”自発音声に対する音素自動ラベリング精度の検証,” 日本音響学会 2001 年春季研究発表会講演論文集, pp.97-98 (2001).
- 3) 菊池英明, ”日本語音声への韻律ラベリング,” 人工知能学会, SIG-SLUD-A003-4, pp.21-24 (2001).
- 4) J. Venditti, ”Japanese ToBI Labelling Guidelines,” OSU Working Papers in Linguistics, Vol.50, pp.127-162 (1997). (Distributed in 1995 as web document; http://www.ling.ohio-state.edu/phonetics/J_ToBI/).
- 5) 菊池英明, 籠宮隆之, 前川喜久雄, 竹内京子, ”自発音声に対する J_ToBI ラベリングの問題点検討,” 日本音響学会講演論文集 (春季), 3-Q-21, pp.383-384 (2001.3).
- 6) 前川喜久雄, 菊池英明, 五十嵐陽介, ”X-JToBI: 自発音声の韻律ラベリングスキーム,” 電子情報通信学会技術報告, NLC2001-71/SP2001-106, pp.25-30 (2001.12).
- 7) 菊池英明, 前川喜久雄, ”日本語自発音声韻律ラベリングスキーム X-JToBI の能力検証,” 人工知能学会研究会資料, SIG-SLUD-A202-06, pp.33-37 (2002).