スペシャル・セッション〔パラ言語・非言語情報の知覚，分析，生成〕

1—10—15 自発音声韻律ラベリングスキーム *X-JToBI* による ラベリング精度の検証 *

○菊池英明 †‡　前川喜久雄 †

† 国立国語研究所　‡ 早稲田大学

1 はじめに

日本語音声の韻律情報のコーディング手法である J_ToBI モデル [1] では、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 [2]。そこで我々は J_ToBI に対して自発音声の韻律ラベリングに必要な拡張を施した新たな韻律ラベリングスキーム X-JToBI(eXtended J_ToBI) を提案した [3]。本稿では、X-JToBI のラベリングスキームとしての能力検証の一環として、ラベル正解率やラベル付与作業者 (以下、ラベラー) 間の一致率などによりラベリング精度を調べた結果を報告する。

2 ラベル出現頻度

X-JToBI では自発音声に特有の現象を記述するためにいくつかの新たなラベルを導入している。ラベリング精度を測る前に、まず始めに実際の自発音声においてこれらがどの程度活用されるかを調べる。以下には、音韻的トーンを表す tone 層と単語間の韻律的区切りの度合を表す break index(以下、BI) 層にわけてラベルの出現頻度を分析した結果を示す。なお、分析の対象としたデータは、現在我々が構築を進めている「日本語話し言葉コーパス」(*Corpus of Spontaneous Japanese*:CSJ)[4] のコア部分に対して一名の作業者が X-JToBI ラベリングを行なったものである。データにおける発話の総時間は模擬講演が 2.75(男性 0.99、女性 1.76) 時間、学会講演が 1.02 時間 (男性のみ) である。

2.1 Tone 層

X-JToBI では句末における局所的なピッチ変化 (boundary pitch movement, 以下 BPM) として従来の"L%H%", "L%HL%" の他に新たに"L%LH%" を導入した。表 1 に示したデータ種別毎の BPM 出現頻度分布から、提案した"L%LH%" が少ないながらもどの種別にも出現していることがわかる。

同様に、tone イベントにおけるピッチ保持を表すために導入したエクステンダー ("<","<") は全ての tone ラベルのうち 1.85%の割合で出現していた。

表 1: BPM を表わす Tone ラベルの出現頻度分布
(() 内は各種別における BPM 総数に対する割合 [%])

tone ラベル	模擬講演		学会講演	
L%+H%	1683	(58.05)	1331	(78.80)
L%+HL%	1121	(38.67)	346	(20.49)
L%+LH%	95	(3.28)	12	(0.71)

2.2 BI 層

表 2 にはデータ種別毎の BI ラベル出現頻度分布を示す。なお、CSJ における韻律ラベリングで使用しないラベル ("1+w" など) は集計の対象から除外した。

この表において量の多少はあっても全てのラベルが利用されていることがわかる。"D" や "F" など、新たに導入したラベルも活用されており (全体の 7.4%)、自発音声特有の現象を表現するのに役立っているといえる。

表 2: BI ラベルの出現頻度分布
(() 内は各種別における BI ラベル総数に対する割合 [%])

BI ラベル	模擬講演		学会講演	
1	31696	(69.53)	8121	(55.73)
1+	42	(0.12)	9	(0.06)
1+p	254	(0.70)	164	(1.13)
2	4071	(7.82)	1333	(9.15)
2+	15	(0.03)	9	(0.06)
2+p	214	(0.41)	67	(0.46)
2+b	466	(0.90)	526	(3.61)
2+pb	48	(0.09)	8	(0.05)
3	7478	(14.37)	2617	(17.96)
D	280	(0.54)	97	(0.67)
D+	31	(0.07)	26	(0.18)
P	19	(0.04)	6	(0.04)
P+	12	(0.02)	5	(0.03)
<F	431	(0.83)	464	(3.18)
F	2426	(4.66)	1111	(7.62)
PB	61	(0.06)	10	(0.07)

3 ラベリング精度

本節では、ラベリング精度を調べた結果を報告する。ラベリング精度としては、暫定的に設定した正解との一致率によって正確性を、複数ラベラー間の一致率によって再現可能性を測定する。なお、ラベラー間一致率の指標としては Cohen が定義した κ[5] を用いる。κ は、観測された一致率を $P(O)$、期待される一致率を $P(E)$ とすると以下の式により得られる。

$$\kappa = \frac{P(O) - P(E)}{1 - P(E)} \quad (1)$$

ここでは、音声的に特徴のある平均約 30 秒程度の講演音声を CSJ から 9 つ抽出して用意し、3 名のラベラーが X-JToBI にしたがって韻律ラベリングを行なった結果を分析の対象とした。なお同じ音声データに対して多数のラベラーが J_ToBI にしたがってラベリングを行なった [2] 結果のうち暫定的に設定した正解に近い上位 3 名のラベルを比較対象として用いた。J_ToBI の 3 名のラベラーは過去にラベリングの経験があり、熟練度は比較的高いといえる。

* Accuracy of Prosodic Labeling of Spontaneous Speech by *X-JToBI*.
By KIKUCHI Hideaki †‡ and MAEKAWA Kikuo † († Nat'l. Inst. for Japanese Language , ‡ Waseda University)

なおラベリングに先立って、語境界および語のアクセント情報を記すword層のラベルを形態論情報に基づいて作成し、BI層の初期値としてword層ラベルと同じ位置にBI=1を与えている。tone層の初期値は与えなかった。

3.1 BI層

まず、BI層について、正解ラベルに対する正解率をラベル種類毎に集計した結果を表3に示す。表中、J_ToBIとX-JToBIの対応するラベルを同一行に示している。この表から、サンプル数の少ない"1+"を除くほぼ全てのラベルにおいて正解率が向上していることがわかる。なお、アクセント句境界にBPMが生じることによって"2"よりも強い区切りになることを示す"2+b"の正解率が、主に対応するJ_ToBIの"2m"の正解率に比べて低くなっている。この原因として、「という」などの引用の形で「と」が高いピッチで発声されるような、アクセント句境界の位置判定が困難になるケースで、熟練したJ_ToBIの3名のラベラーの方がパターンとして処理できていることがあげられる。こうしたケースに対しては特に基準を明確にしたうえでラベラーの訓練を行なう必要がある。

表3: BIラベルの正解率
(() 内は正解ラベルにおける総出現数)

J_ToBI			X-JToBI		
ラベル	正解率		ラベル	正解率	
1	91.3	(593)	1	94.9	(531)
2	74.0	(123)	2	74.4	(112)
3	70.5	(182)	3	75.1	(181)
2-	33.3	(1)	1+	0.0	(1)
1p	47.9	(16)	1+p	81.5	(9)
3-	–	(0)	2+	–	(0)
2m	80.5	(29)	2+b	66.7	(30)
2p	27.3	(11)	2+p	33.3	(8)
3m	0.0	(1)	2+pb	33.3	(4)
			D	83.3	(4)
			D+	44.4	(3)
—	—	—	<F	76.2	(7)
			F	91.0	(63)
			PB	33.3	(2)
全体	83.2	(956)	全体	86.1	(956)

次に、ラベラー間一致率を調べたところ、J_ToBIにおいて$\kappa = 0.64$ ($P(O) = 0.78$, $P(E) = 0.40$)であり、X-JToBIにおいては$\kappa = 0.73$ ($P(O) = 0.83$, $P(E) = 0.37$)であった。ラベルの種類を増やしたにもかかわらず一致率が向上していることから、本スキーム設計の有効性がうかがえる。

なお、X-JToBIの正解として"D"、"F"が付与されている箇所を除いてラベラー間一致率を求めたところ、J_ToBI、X-JToBIにおいてそれぞれ$\kappa = 0.69$、$\kappa = 0.71$と差が小さくなることから、特に"D"と"F"の効果が大きいといえる。

3.2 Tone層

次に、tone層のうちBPMのラベリング精度を調べた。まず、正解に対してアクセント句境界が一致する箇所のみを対象に句末境界音調毎に正解率を調べたところ、一致箇所の数と正解率はそれぞれJ_ToBIが251、84.1%、X-JToBIが635、87.7%であった。

先に、自発音声に対するtoneラベリングの誤り要因を分析したところ、アクセントの知覚誤りに次いでBPMの知覚誤りが高い割合を示していた[2]。我々は、この問題がBPM判断基準の不明確さに起因すると考えて聴覚による判断を物理的な特徴に基づく判断よりも優先させることを明確にし、そのうえで物理的イベントにラベルを対応づけることによってBPM知覚への注意を促すよう基準を設けたが、上述の結果はその効果を実証しているといえる。

次に、BPMのラベリングにおけるラベラー間一致率を調べたところ、J_ToBIにおいて$\kappa = 0.41$ ($P(O) = 0.53$,$P(E) = 0.19$)、X-JToBIにおいて$\kappa = 0.61$ ($P(O) = 0.68$,$P(E) = 0.18$)であった。このことから、本スキームによってBPMのラベリングにおける正確性および再現可能性が向上することがわかった。

なお、句頭の上昇音調を表す"H-"については、正解率においてJ_ToBIの77.3%に対してX-JToBIは74.7%と若干下降しているが、ラベラー間一致率ではJ_ToBIの$\kappa = 0.48$に対してX-JToBIは$\kappa = 0.74$と大幅に向上している。また、アクセントについても正解率はJ_ToBI、X-JToBIで83.8%、87.9%と差がないがラベラー間一致率ではJ_ToBIの$\kappa = 0.45$に対してX-JToBIは$\kappa = 0.61$と大幅に向上している。

4 おわりに

本稿では、日本語自発音声の韻律ラベリングスキームX-JToBIの能力検証の一環として、ラベル出現頻度を調べたうえでラベリング精度を調べた。その結果、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。

本稿では主にJ_ToBIによるラベリング結果との比較によりX-JToBIの有効性を議論したが、X-JToBIのラベリング結果について、観測された一致率($P(O)$)を見るとtoneラベルのうちのアクセントとBPMにおいてはまだ充分な精度とは言えない。今後、さらなる工夫が必要であろう。

参考文献

[1] Venditti, J., "Japanese ToBI Labelling Guidelines," Manuscript. Ohio State University, USA., 1995.

[2] 菊池他, "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, pp.383-384, 3-Q-21, 2001.3.

[3] 前川他, "日本語自発音声の韻律ラベリング体系:X-JToBI," 日本音響学会講演論文集, pp.313-314, 3-10-6, 2002.3.

[4] 前川他, "日本語話し言葉コーパスの設計," 音声研究, vol.4, no.2, pp.51-61, 2000.8.

[5] Cohen, J., "A Coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, Vol.20, No.1, pp.37-46, 1960.