

X-JToBI: 自発音声の韻律ラベリングスキーム

前川喜久雄[†]・菊池英明[†]・五十嵐陽介[‡]

[†]国立国語研究所 研究開発部門 [‡]東京外国語大学大学院地域文化研究科

要旨 本稿では日本語の韻律ラベリングスキームである J_ToBI を拡張した新スキーム、X-JToBI(eXtended J_ToBI)を提案する。新スキームでは分節音ラベリングが実施されていることを前提として、J_ToBIの問題点であった tone ラベル位置と物理的イベントの不一致を解消した。さらに自発音声の韻律ラベリングを実施するために必要とされる多くの拡張を施した。この拡張には、BPM インベントリの拡張と表示方法の変更、BI ラベルの詳細化、フィルター用ラベルの提案、卓立のラベリング法などが含まれる。

X-JToBI: An Intonation Labeling Scheme for Spontaneous Japanese

Kikuo Maekawa[†], Hideaki Kikuchi[†] and Yosuke Igarashi[‡]

[†]Dept. Language Research, National Institute for Japanese Language

[‡]Graduate School of Area and Culture Studies, Tokyo University of Foreign Studies

ABSTRACT Outline of the X-JToBI intonation labeling scheme, the extended version of the J_ToBI, is presented. The main motive for the extension being consisted in its application for spontaneous speech, the following extensions were introduced: 1) Exact match between the time-stamp of tone labels and the timing of physical events, 2) Enlargement of the inventory of Boundary Pitch Movement, 3) Extension and ramification of the usage of Boundary Indices, and 4) Proposal for the labeling of filled-pause and non-lexical penult prominence.

1. はじめに

J_ToBI は音韻論に立脚した日本語の韻律ラベリングスキームであり、提案以来幅広い支持を受けてきている(Venditti, 1995=1997)。

しかしながら、現在我々が構築をすすめている『日本語話し言葉コーパス』(前川他 2000、以下CSJと略称)に代表される自発音声コーパスに対して現行のJ_ToBIを適用しようとするると種々の困難に逢着する(菊池 他 2001)。

以下に提案する X-JToBI(eXtended J_ToBI)は、J_ToBI を自発音声にも適用可能なように拡張する試みであるが、拡張の動機の中には J_ToBI 自体の問題点の解消も含まれている。以下、2節では J_ToBI 固有の問題点に触れ、その改良方針を示す。続いて3節では自発音声に対処するために導入した拡張要素について説明する。最後に4節で音韻論上の問題点に触れる。

2. J_ToBI 固有の問題

2. 1. 1 時間情報

現行の J_ToBI に対して多くのユーザーが感じる不満のうち最大のものは、tone 情報に関連した時間情報の取り扱いである。J_ToBI では tone および break indices(BI)のラベルに音声信号に対する時間情報(time stamp)を付与して表示するが、その際以下の慣

行に従わねばならない。

- 1) 句頭の上昇の頂点を示す H- ラベルは音韻論的にそれが属するモーラないし音節の時間区分内部に位置させる。
- 2) 語彙アクセントを示す H*+L ラベルは、音韻論的にそれが属するモーラないし音節の時間区分内部に位置させる。
- 3) 韻律句(intonation phrase)末に生じる境界音調(boundary tone)のラベルは、それが所属する BI の終端に同期させる。
- 4) BI はそれが属する語の終端に同期させる。

このうち 1)-3)には問題がある。1)と 2)に関して問題となるのは、イントネーションの物理的関連量としての Fo の頂点(ないし屈曲点)が、対応する音韻論的時間区分の外部にはみだしている場合である。その場合 J_ToBI では tone ラベルに加えて“>”(early peak)や“<”(late peak)を用いてピークの物理的位置を表示する。

この慣行は、tone ラベルのなかに部分的に分節音の時間区分情報を導入するものであり、しかもその情報は十分に精密なものではない(例えば“>”や“<”がどのモーラの時間区分に位置しているかを知ることができない)点に問題がある。

この問題は J_ToBI が word-tier に表示される語単位の区分情報を除いては分節音持続時間長の情報を保有しないために生じる問題である。

3)については上述の問題に加えて、複合境界音調に関する問題がある。J_ToBI では L%HL%のように複数の頂点を有する複合境界音調(complex boundary tone)は、その全体に対するラベルを BI 位置に付与する慣行である。そのためたとえ“>”ラベル等を用いても一部の頂点の物理的生起位置を記述することができない。そのため複合境界音調ラベルは本質的に分類ラベルとしての役割しか果たしていない。

2. 1. 2 解決策

上記 1)-3)に関する根本的な解決策は分節音に関する組織的なラベリングを実施することである。分節音ラベリングが種々のポーズを含めて発話全体に施されていれば、word-tier に記述されている言語情報(語彙アクセントの音韻論的位置や単語のモーラ・音節構造の情報)から、アクセントおよび H-のオーソドックスな実現期待位置を決定することができる。従って境界音調も含めてすべての tone ラベルはその物理的な実現位置に付与すればよい。H は Fo の局所的最高値ないし屈曲点、L は局所的最低値ないし屈曲点をラベル位置とする(ただし複合境界音調を単純に分解してしまうと検索上の困難が生じる。次節参照)。図 1 以下の tone ラベルはすべてこの方式で付与している。

分節音ラベリングには精粗様々なレベルが考えられる。最も粗い方法は音声認識エンジンを利用した自動ラベリング結果をそのまま利用する方法である。この場合ラベリングのコストはほとんど増加しないがラベリング精度を向上させようとするればそれだけコストも増加する。CSJ では HTK(Young et al.,1995)の出力を初期値として手作業による修正を施している。

3. 自発音声韻律ラベリングのための拡張

本節では CSJ のラベリング過程で導入した拡張を、tone、BI、フィルターに分けて説明する。

3. 1 Tone ラベル

3. 1. 1 句末境界音調

句末境界音調については 1)境界音調インベントリを追加し、2)複合境界音調の表示方法を変更した。1)については、従来から認定されている L%, L%H%, L%HL%に加えて、L%LH%を認定することにした。この境界音調は低く始まり、一定時間低さを持続させた後に上昇に転ずるタイプのもので、反問や強意などのパラ言語情報を伝える際に多用されるものである。川上(1963)による文末上昇の分類では第三種の上昇(反問の上昇;図 2 参照)に該当する。

次に表示方法の変更であるが、先に 2.1.2 で述べたように、我々は分節音ラベリングを前提として各 tone を物理的イベントに対応させることを原則としている。そのため例えば L%HL%は、L%, H, L% の三つのイベントに分解されるのであるが、これらの記号をそのままラベルとして利用すると検索時に支障をきたす。そのため、tone ラベルとしては L%と HL%のふたつを付与し、Fo の頂点には、pH という補助ラベルを付与することにした。この補助記号のことをポインターと呼ぶ。ポインターには pH と pL

の 2 種類がある。

表 1 に句末境界音調の一覧とそのラベルを示し、図にラベリング例を示す。有核アクセント句「名古屋で」を想定した模式図である。

Table 1. Inventory of phrase-final boundary tones and the corresponding X-JToBI labels.

Type	Labels
L%	L%
L%H%	L%, H%
L%HL%	L%, pH, HL%
L%LH%	L%, pL, LH%

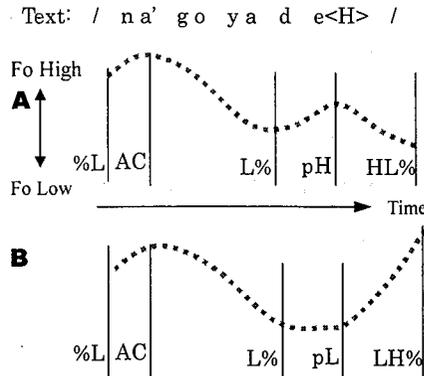


Figure 1. Schematic examples of the tone labeling of phrase-final complex boundary tones. Pointers (pH and pL) are used to denote the time-points of H and L of HL% and LH% respectively. AC stands for accent.

/ n a' g o y a d e<H> /

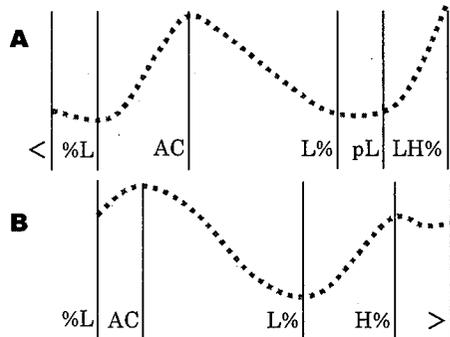


Figure 2. Examples of the usage of extenders. At the utterance initial (top) and utterance final positions (bottom). The rendition of A is a typical contour of “suspicion/disbelief”.

3. 1. 2 句頭境界音調

J_ToBI における句頭音調は %L 一種である(ただし 4.3 節参照)。しかし、ある種のパラ言語情報の伝達にはアクセント句頭のイントネーションが関与することが知られている(川上 1956;前川・北川,投稿中)。就中、川上が「遅上り型」と呼ぶタイプ、つまりアクセント句頭におけるピッチの上昇タイミングが第

2 モーラ以降に位置するタイプが重要である。このイントネーションの情報をラベル化するためには、アクセント句頭の局所的最低 Fo 値を示す%Lに加えて、低ピッチが保持されている区間を示すラベルが必要である。我々は図2に示すように、このタイミングを記号">"ないし"<"によって表示することとし、この記号をエクステンダーと呼んでいる。図2Aでは句頭における最低 Fo 値の時刻に%Lが付与されており、エクステンダーは低ピッチ区間の冒頭に付与されている。このケースのようにエクステンダーが対応する tone の左側に位置する場合は記号"<"を、反対の位置関係の場合は記号">"を用いる。

エクステンダーは句末境界音調の表示に利用されることもある。図2Bがその例である。ここではL%H%で記述される句末音調の末尾が主観的には高いピッチを保ったまま引き延ばされている。ここでH%はFoの最高値時刻に付与されており、エクステンダーは高ピッチの延長の終端を示している(3.3も参照)。

なおJ_ToBIでは%LにStrong vs. Weakの二種類を認めているが、X-JToBIではこれを廃止した。

3. 1. 3 Foの異常と欠損

現実のデータではtoneが存在すると期待される時刻のFoが抽出誤りによって異常値となったり、母音無声化等によって欠損したりすることがある。その場合TをtoneとしてTx(L%x, H-x等)によってFoの欠損を、またT?によってFo値の異常性を示す。

3. 2 Break Indices (BI)

J_ToBIの運用ではBIの判定に迷うケースが少なからず生じる。この場合に備えてJ_ToBIには“p”, “-”, “m”の補助記号が準備されているが、これらを具体的にどのようなケースに適用するかの説明が不足しているために作業に支障をきたすことが多い。自発音声のラベリングではこの困難が極端に増幅される。X-JToBIではBIのインベントリを拡張するとともに、補助記号を整理して個々の用法を詳細化することに努めた。

表2にX-JToBIにおけるBIおよび補助記号の一覧を示す。用法中のBPMはboundary pitch movementの略であり、句末境界音調のうちピッチの上昇下降が認められるケース、つまりL%が単独で生じるケース以外のすべての音調を意味している。備考中のNRはNot Recommended(推奨せず)の意味。同じくOPはOptionalの意である。OP要素を利用する場合はコーパス全体に一貫して利用することが必須である。

A 整数0-3の用法は従来どおりであるが、0はオプションとした。これは従来0が適用される対象が語の融合(Ex.私は>ワタシヤ)に限られており、本来的に形態論レベルの情報しか提供していないと考えられるからである。CSJのラベリングでも0は利用していない。

整数値1-3に関しては、補助記号“+”によって連続する整数の中間値の表現を許容する。また、中間値が生じた原因を表示する補助記号として“w”, “p”, “b”を用意し、+に後続させる。1+, 2+など原因を明示しない用法も許容するが推奨しない。

B “1+w”は複合語中の韻律境界を表示する。例えば複合名詞である「名古屋大学医学部」は全体が1ア

クセント句となることもあるが、ふたつのアクセント句として実現されることもある。その場合X-JToBIでは「ナゴヤダイガク(1+w)イガク」のように複合語中の境界に“1+w”を付与する(J_ToBIではBIは2となる)。“1+w”はOPであるが、CSJでは日本語に頻発するこの種の語中韻律境界の性格を検討するためにラベルを付与している。なお「名古屋大学文学部」の下線文にフォーカスが置かれることによって、後半要素のピッチレンジがリセットされている場合には(J_ToBIと同様に)BI=3を付与する。

Table 2. Inventory of BI labels used in X-JToBI

Label	Usage	Remarks
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+	J_ToBI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	

/ayama' ri <pz> t o narima'su/

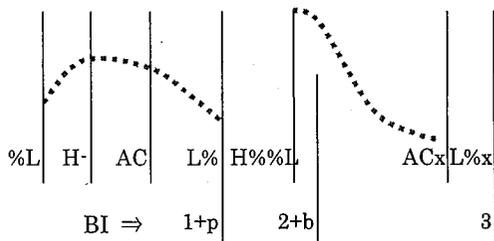


Figure 3. Example of the usage of “1+p” and “2+b”. <pz> means perceptible pause. Fo values corresponding to “AC” and “L%” at the phrase end are lacking due to vowel devoicing.

C 図3は文節を構成する実質語と形式語の間にポーズ<pz>が介入している例である。実質語末のBIラベルは“1+p”になる。この種の発音は話者によっては講演や講義に頻出する。その際、この図のように後続する助詞にH%境界音調がおかれ、また後続アクセント句との間にポーズが介在しないことが多い。

D 図3ではアクセント句境界に“2+b”が与えられている。“2+b”はアクセント句末にBPMが存在するが、ポーズは後続せず、なおかつ後続するアクセント句のピッチレンジがリセットされていない(先行句が有核ならdownstepが継続している)ケースに付与するBIである。

J_ToBIでは補助記号“m”(mismatch)を適用するケースであるが、“2m”と“3m”に解釈が分れるので、新ラベルによって統一した。

“2+bp”は、アクセント句境界にBPMが存在し、ポーズも後続する(以上はBI=3の特徴)が、後続アクセント句のピッチレンジがリセットされていない(downstepが持続する)場合に利用する。

図3で“2+b”の境界に付与されている“H%L”は、先行アクセント句末の境界音調H%と後続アクセント句頭の%Lが同一時刻に付与されることを示している。“2+b”境界に有声音で始まるアクセント句が後続する場合、ピッチは後続アクセント句全体を通じてただらかに下降を続けることが多く、%L位置はその先頭と考えられるからである(4.2節参照)。

E “2+p”はアクセント句末にBPMが存在せず、ポーズが後続し、アクセント句間でピッチレンジのリセットが認められない場合に用いられる(図4)。J_ToBIのマニュアルにはアクセント句にポーズが後続することがBI=3の要件であるかに読める記述があるためにラベリングにゆれが生じることが多い。“2+p”はその点を明確化している。

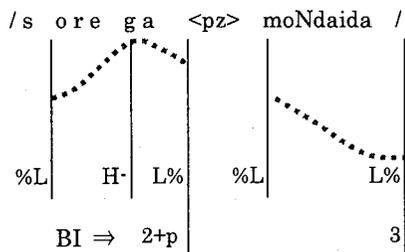


Figure 4. Example of the usage of “2+p”.

F “D”と“P”(共に大文字)は従来の整数値の替りに利用するBIである。まず“D”は語の断片が存在することによって生じる韻律境界を示すBIである。例えば図5Aのように「名古屋」を意図して「ナゴ<pz>ナゴヤ」と発音した場合、冒頭2モーラが語断片となる。その際、語断片と後続語の間にポーズが存在する、後続語に句頭上昇が認められる等、韻律上の境界が存在するならば語断片末のBIとして“D”を用いる(韻律境界が感じられなければBIは何も付与しない)。

語の後半が語断片となることもある。例えば「藤崎先生が」を意図した話者が誤って「藤村」と発音

した直後に語の後半要素だけを言い直すと、「フジムラサキセンセーガ」という発話が生じる。この場合、図5BのようにBIをふる。ここで“D+”が用いられているのは、その境界の後ろでピッチレンジがリセット(拡大)されていることを示すためである(ただし+はオプション、表2参照)。

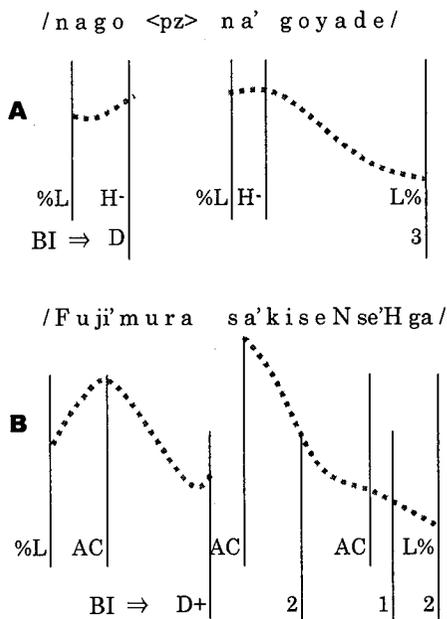


Figure 5. Example of the usage of “D” and “D+”.

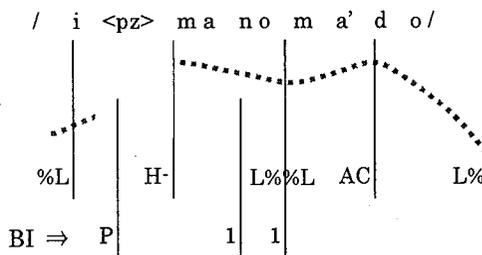


Figure 6. Example of the usage of “P” showing word-internal pause.

G “P”は語(形態素)の内部にポーズが位置することによって生じた韻律境界に付与されるラベルである。図6はポーズの前後でピッチレンジに変化がない場合を示しているが、ポーズの後でピッチレンジが拡大されていれば“P+”を付与することができる(オプション)。

残る“<F”と“F”および“PB”については3.3節および4.2節でそれぞれ解説する。

3.3 フィラー

CSJには大量のフィラーが含まれており、その分析は“D”や“P”の分析と並んで自発音声研究の重

要な課題になると思われる。J ToBI ではフィラー (filled pause) のラベリング方法が規定されていない。あるいは通常のラベリング手法がそのまま適用できると想定されているのかもしれないが、実際にはフィラーには固有の韻律特徴が存在しており、それに特化したラベリングが要請される。

C S J の転記テキストでは「ワタシガー、コレオー」のように句末の母音が延長されていれば「母音の非語彙的な延長」を示す<H>タグが付与されている。また「エー」「アノー」「マ」「ア」の様に語彙的に独立した要素に対してはフィラー候補のゆるやかな一覧表を作成し、表に含まれる要素を (F) タグによって囲んでいる(小磯他 2001)。韻律ラベリングにおいては、これらの中から韻律的に特別な処理を必要とするもの(韻律的フィラー)とそうでないものとを判別しなければならない。

まず句末母音が延長されるタイプは BPM として処理する。その際、図 2 B のようにエクステンダーを利用することが必要となることもある。

次に語彙的に独立したフィラー候補は、主観的長さが 1 モーラのものとは 2 モーラ以上のものにおいて処理する。1 モーラの候補 (「ア」「エ」「マ」等) はすべて韻律的フィラーとみなす。2 モーラ以上のフィラーは 1) 句頭の上昇が認められず、2) アクセント核類似の局所的なピッチ下降も認められないもの、つまり全体を通してピッチが平坦なものだけを韻律的フィラーと認定する。上記 1), 2) のいずれかひとつにでも該当すれば通常のアクセント句としてラベリングする。

図 7 に韻律的フィラーのラベリング例を示した。韻律的フィラーが通常のアクセント句に直続している場合、フィラー始端に「<F」、終端に「F」という BI を与える。韻律的フィラーの前後にポーズが生じている場合、および韻律的フィラーに通常のアクセント句が直続している場合には、韻律的フィラーの終端にだけ「F」を与える。

韻律的フィラーのピッチは平坦であるので、tone はその全体に対して「高」か「低」かだけを判定し、それぞれ「FH」、「FL」ラベルで表示する。通常のアクセント句に韻律的フィラーが直続する場合、アクセント句末尾 tone の水準がフィラーの高さを決定することが普通であるが、その場合もラベルを省略することはしない。

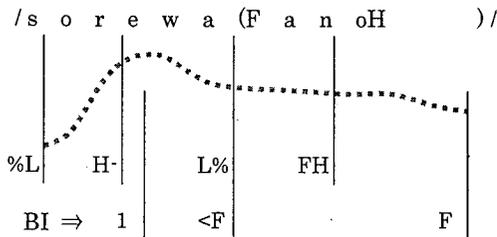


Figure 7. Example labeling of filled pauses. “<F” and “F” are special BI for filled pauses. Tone labels “FH” and “FL” are given to the mid point of a filled-pause that is embraced by the tag (F).

4. 音韻理論上の問題

J ToBI は音韻理論を背景にもつラベリングスキームである。この基本的性格は X-JToBI にも継承されている。

実際に X-JToBI を自発音声データに適用してみると、理論の根幹にかかわる種類の現象に出会うことがある。本節ではその種の問題を三例検討する。

4. 1 卓立 (Penult Non-Lexical Prominence)

自発音声では「データが」が「デー」「タ」「ガ」のように、同じく「モーラが」が「モ」「ー」「ラ」「ガ」のように発音されることがある (↑と↓でピッチの局所的な下降と上昇を示す)。つまり語彙的なアクセントの他に、句末から二番目 (penult) のモーラのピッチが選択的に高められる現象である。

この現象は国語学の世界では早くから研究されており、「卓立」「プロミネンス」などと呼ばれている (川上 1957, 大石 1959)。国語学では penult 以外の位置にも卓立を認めるが、それらは J ToBI の枠組みで言えばピッチレンジの局所的な拡大ないしは句頭の上昇の遅れによって記述が可能 (従って X-JToBI でも記述可能) と思われる。それに対して penult 位置の卓立は特別な処理が必要とされる(谷口 2001)。

X-JToBI の枠組みで penult 位置のプロミネンスを解釈すると、そこに局所的なピッチ上昇として実現する特殊な tone が存在していると考えられる解釈と、句末境界音調の変異形とみなす解釈との二通りが可能である。現在 C S J の韻律ラベリングでは tone のインベントリを拡張する必要がないことを重視して第二の解釈を採用している。

図 8 に penult 卓立のラベリング例を示す。通常の句末境界音調 L%HL%に用いるラベルが用いられているが、各ラベルとモーラとの時間関係が変化しており、L%は「モー」末尾付近の Fo 最低値に、pH は卓立の頂点位置に、それぞれ配置されている。

Penult 卓立をこのようにラベリングすると、ただ tone-tier のラベルを検索するだけでは卓立の有無を知ることができない。分節音ラベルを参照すれば検索可能であるが、複雑である。Penult 卓立の存在を miscellaneous-tier に記入しておくことが望ましい。C S J では “PNLP” というラベルを用いている。

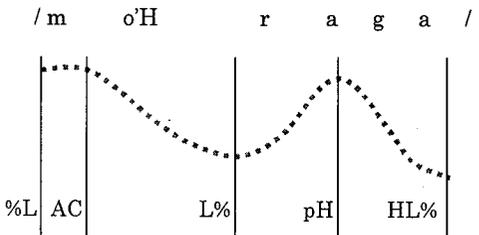


Figure 8. Example tone labeling of Penult Non-Lexical Prominence. PNLN is regarded as a variant of L%HL% boundary tone.

4. 2 Parasitic Boundary

“PB”は俗に「半クエスション」と呼ばれるイントネーション等に絡んで稀に適用される BI ラベルであ

る。図9は「農業の衰退が」という発話の例であるが、「農業」の末尾に半クエスチョンによる上昇が生じたうえ、後続する格助詞にも句末境界音調(H%)が生じている。境界音調の継起はJ_ToBIでも、その理論的基礎をなす Pierrehumbert & Beckman (1988)でも全く想定されていない現象である。

これに対処するためイントネーション文法を改定して継起を許容することも考えられる。しかしそれでは文法の生成力が不必要に拡張されてしまう。図9に類する現象は現在のところ半クエスチョンと句末上昇を有する句が引用された場合(例「本当?と言いました」)の二種類しか観察されていないこと、および半クエスチョンが日本語の韻律現象として不安定であること(多くの日本語話者に違和感を抱かせる)の二点を考慮すると、J_ToBI 文法の例外として処理することが望ましいと考えられる。

“PB”は parasitic boundary の略であり、文法的に許容されない韻律境界が寄生的に発生していることを示すラベルである。現在のところ“PB”を適用するのは半クエスチョンに句末境界音調が後続する場合と句末音調を有する句が引用された場合のみである。後ろに句末境界音調を伴わない半クエスチョンや、句末音調なしの引用は通常の句末境界音調(L%H%)としてラベリングすればよい。

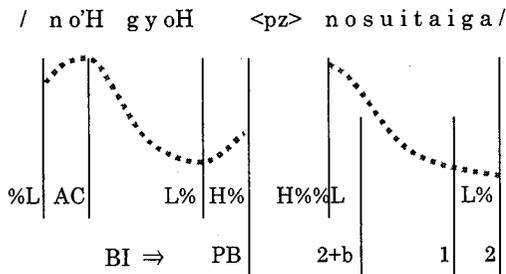


Figure 9. Example of “PB”. This label is applied to ungrammatical succession of BPMS and demarcates the end of the first one. In this case, the rise at the end of /noHgyoH/ is so-called ‘quasi-question’.

4. 3 句頭境界音調の問題

3.1.2で指摘したように、J_ToBI が認定する句頭境界音調は%L一種類だけである。しかし実際のデータ中には、アクセント句頭に%Hを認定したくなるケースがある。

例えば談話の冒頭がいきなり「トユーワケデ」あるいは「デダスネー」などと始まることがある。この種の発話のピッチは高い値で始まり急激かつ単調に下降しつづけることが多い。ちょうど図3ないし図9に示したイントネーションの右半分が、独立して用いられる形である。

現在のところ X-JToBI には“%H”を導入しておらず、上の発話に対しては発話冒頭に“(H%)%L”というラベルを付与している。上に例としたふたつの発話はいずれも、「～トユーワケデ」あるいは「～デダスネー」のように、発話に先行する句(～)が句末の助詞を残して省略された形をしている。“(H%)%L”は、省略された句末の境界音調“(H%)”が現実の発話

の冒頭に痕跡として残存しているという解釈に基づくラベルである。音韻論的には“(H%)%L”は通常の“%L”の条件異音と解釈することができるので、J_ToBIと同様、X-JToBIにおいても句頭境界音調は音韻論的には一種類だけである。今後もし“%L”の条件異音とみなすことが困難なケースが発見されれば、あらためて“%H”の導入を検討することになるだろう。

5. おわりに

本稿では自発音声への適用を目的としたJ_ToBIの拡張について報告した。自発音声の韻律は想像を越えて多様であり、例えば“PB”によって処理することにしたイントネーションなどは、実際にそれを耳にするまでは想像することができなかった。今後も新しい拡張を導入する必要がある可能性は否定できない。それについては適宜報告することにした。

最後に、現在までのところ X-JToBI によるラベリング結果はJ_ToBIラベルに自動変換することが可能であることを付言しておく。

謝辞: J_ToBI による自発音声ラベリングの問題点を把握するにあたっては、2000年11月に国立国語研究所で開催された『自発音声韻律ラベリングワークショップ』における議論が有益でした。参加いただいた皆様に深く感謝いたします。Jennifer Venditti 氏からはこのワークショップ以外にも随時コメントをいただきました。X-JToBI の仕様の一部には同氏のアイデアが反映されていることを記して感謝します。

文献

- 大石初太郎(1959).プロミネンスについて. 国立国語研究所 論集 I ことばの研究.
- 川上泰(1956).文頭のイントネーション. 国語学, 25, 21-30.
- 川上泰(1957).東京語の卓立強調の音声. 国語研究, 6, 21-31.
- 川上泰(1963).文末などの上昇調について. 国語研究, 16, 25-46.
- 菊池英明 他 (2001). 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), 1, 383-384.
- 小磯花絵 他 (2001). 「日本語話し言葉コーパス」における書き起こしの方法とその基準について. 日本語科学, 9, 43-58.
- 谷口美希(2001).句末上昇について—J_ToBIにおけるH%—. 人工知能学会研究会資料(SIG-SLUD-A101-4), 13-18.
- 前川喜久雄他(2000).「日本語話し言葉コーパス」の設計. 音声研究, 4-2, 51-61.
- 前川喜久雄・北川智利(投稿中). 音声はパラ言語情報をいかに伝えるか. 認知科学.
- Pierrehumbert, J. & M. Beckman (1988). *Japanese Tone Structure*. MIT Press.
- Venditti, J. (1997). Japanese ToBI Labelling Guidelines. OSU Working Papers in Linguistics, 50, 127-162. (Distributed in 1995 as web document; http://www.ling.ohio-state.edu/phonetics/J_ToBI/).
- Young, S. et al. (1995). *The HTK Book*. Entropic Research Laboratory.