

『日本語話し言葉コーパス』の分節音ラベリング

Version 1.1 (2011年10月10日)

前川喜久雄*・菊池英明†・藤本雅子*・米山聖子‡・西川賢哉**

(*国立国語研究所 †早稲田大学 ‡大東文化大学 **理化学研究所)

目次

1. 本文書の内容	2
2. 分節音ラベリングの方針	2
3. 分節音ラベル	2
4. ラベリング方式	4
4. 1 分節音ファイルの形式と注意点	4
4. 2 ラベリングの基準	6
A) 発話単位に関するラベル	6
B) 子音ラベルおよび関連する補助ラベル	6
C) 母音ラベルおよび関連する補助ラベル	8
D) その他の補助ラベル	8
5. 利用上の注意	8
5. 1 ラベルの融合	8
5. 2 <cl>について	9
5. 3 転記テキスト中の<H>, <Q>について	10
6. ラベリング作業の実施方法	10
7. Phone について	10
8. XML 文書および音響モデルにおける「音素」について	11
参考文献	11

更新履歴

Version 1.1

- ・ 促音“Q”に関する記述を修正
- ・ Phone に関する記述（7節）を追加

1. 本文書の内容

本文書は、『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese: 以下 CSJ)における音声ラベリングの仕様を解説したものである。CSJ では、「コア」に含まれる約 50 万語 (短単位)、およそ 44 時間の音声に対して、分節音ラベルとイントネーションラベルを付与しているが、本文書ではこのうち分節音ラベルだけを扱う。イントネーションラベルについては [intonation.pdf](#) 参照。

分節音 (およびイントネーション) ラベルの情報はテキストファイル形式で提供される他、CSJ の XML 文書にも格納されるが、本文書では XML 文書については解説していない。XML 文書については [xml.pdf](#) 参照。

また、本文書 version 1.0 では一部に古い仕様が解説されていたため、CSJ 第一刷の段階から、本文書とラベルデータとの間に齟齬が生じていた。CSJ 第三刷作成 (2011 年) にあたり、分節音ラベリングの最終の仕様を本文書に反映し、それに伴い version を 1.1 に変更した。特に大きな修正は 4.2 「促音」に加えられている。

2. 分節音ラベリングの方針

分節音ラベリングとは、音声信号を、それを構成すると考えられる子音、母音等の構成要素に分解して、当該要素を表現するラベルと時刻を確定する作業である。構成要素は音声学のレベルでも音韻論のレベルでも考案できるが、CSJ のラベルセットは、日本語の音素ラベルに音声学的なラベルを一部追加したものとなっている。これは、現代日本語に観察される分節音の音声変異のうち、音声学的に興味深く、また一貫したラベリングが可能であるものを記録するためである。その結果、CSJ の分節音ラベルは、一般音声学でいう精密表記と簡略表記の中間的なレベルのラベルセットとなっている。以下では分節音ラベルに言及する場合、“a” , “b” のように二重引用符で囲う。

3. 分節音ラベル

表 1～3 に CSJ で用いた分節音ラベルの一覧を示す。表 1 は母音のラベル、表 2 は子音のラベル、表 3 は補助ラベルである。これらのラベルはすべて通常の ASCII 文字の組み合わせで表現されており、小文字と大文字を区別する。なお、表 2 の下部に位置している網掛けを施したラベルは、ラベルセットとしては定義したが、現在は利用していないラベルである。本文書の付録として、現代日本語におけるモーラと CSJ 分節音ラベルの対応表を掲載している。

表 1 CSJ の母音ラベル

ラベル	対 象	ラベル	対 象	ラベル	対 象
a	ア	aH	アの長母音*	A	無声化したア
i	イ	iH	イの長母音	I	無声化したイ
u	ウ	uH	ウの長母音	U	無声化したウ
e	エ	eH	エの長母音	E	無声化したエ
o	オ	oH	オの長母音	O	無声化したオ

*ラベル “H” は長母音の第二要素を表す。転記テキストにおける<H>との相違については、本文書の 5.3 参照。

表2 CSJの子音ラベル

ラベル	対 象
k	カ行子音 (軟口蓋無声破裂音)
g	ガ行子音 (軟口蓋有声破裂音)
s	サ行のうち「サ、ス、セ、ソ」の子音 (歯茎無声摩擦音)
z	ザ行子音 (歯茎有声摩擦音ないし破擦音) → 5.2 参照
t	タ行のうち「タ、テ、ト」の子音 (歯茎無声破裂音)
c	タ行のうち「ツ」の子音 (歯茎無声破擦音)
d	ダ行のうち「ダ、デ、ド」の子音 (歯茎有声破裂音)
n	ナ行子音 (歯茎鼻音)
h	ハ行のうち「ハ、ヘ、ホ」の子音 (声門無声摩擦音)
F	ハ行のうち「フ」の子音 (両唇無声摩擦音)、IPA ならば[ɸ]
b	バ行子音 (両唇有声破裂音)
p	パ行子音 (両唇無声破裂音)
m	マ行子音 (両唇鼻音)
y	ヤ行 (拗音) 子音 (硬口蓋接近音)
r	ラ行子音 (歯茎弾き音)
w	ワ行子音 (両唇軟口蓋接近音)
v	ヴァ行子音 (唇歯有声摩擦音)
kw	「クワ」の子音、いわゆる合拗音、IPA ならば[k ^w]
gw	「グワ」の子音、いわゆる合拗音、IPA ならば[g ^w]
kj	母音イの前で口蓋化した k : 「キ」の子音
gj	母音イの前で口蓋化した g : 「ギ」の子音
sj	母音イの前で口蓋化した s : 「シ」の子音、IPA ならば[ʃ]
zj	母音イの前で口蓋化した z : 「ジ」の子音、IPA ならば[ʒ]ないし[dʒ]
cj	母音イの前で口蓋化した c : 「チ」の子音、IPA ならば[tʃ]
nj	母音イの前で口蓋化した n : 「ニ」の子音、IPA ならば[n̟]
hj	母音イの前で口蓋化した h : 「ヒ」の子音、IPA ならば[h̟]
ky	カ行拗音の子音
gy	ガ行拗音の子音
sy	サ行拗音の子音
zy	ザ行拗音の子音
cy	タ行拗音の子音 (「チャ」「チュ」「チョ」の子音)
ty	「テャ」「テュ」「テョ」の子音、IPA ならば[tj]
dy	「デャ」「デュ」「デョ」の子音、IPA ならば[dj]
ny	ナ行拗音の子音
hy	ハ行拗音の子音
Fy	「フャ」「フュ」「フョ」の子音、IPA ならば[ɸj]
by	バ行拗音の子音
py	パ行拗音の子音
my	マ行拗音の子音
ry	ラ行拗音の子音
N	撥音
Q	促音
G	ガ行子音 (軟口蓋有声摩擦音) *
Gj	母音イの前で口蓋化した G
Gy	ガ行拗音の子音 (軟口蓋有声摩擦音)
@	ガ行子音 (軟口蓋鼻音)
@j	母音イの前で口蓋化した@
@y	ガ行拗音の子音 (鼻音)

*網掛け部分は現在の CSJ では利用していない

表3 CSJの補助ラベル

ラベル	対象
<cl>	破裂音ないし破擦音中の閉鎖区間の終端（バースト位置）
<pz>	発話中のポーズ（200ms 未満）
<uv>	声帯振動終了後に母音フォルマントが持続している場合、その終端を示す
<sv>	母音終了後に声帯振動が持続している場合、その終端を示す
<fr>	ボーカル・フライが言語音と重畳せずに生じている区間の終端に付与
<fv>	母音であるが、音素としての属性が分類不能のもの（母音不確定音）
<?>	子音であるが、音素としての属性が分類不能のもの（子音不確定音）
<N>	種々の雑音
	呼吸音
#	発話単位の先頭に付与

4. ラベリング方式

4. 1 分節音ファイルの形式と注意点

分節音ラベルを記録したファイルは、PLABEL ディレクトリに格納された拡張子が .seg のファイルである。これを以下では分節音ファイルと呼ぶ。分節音ファイルは、Entropic 社の waves 形式のテキストファイルである。ファイルの先頭には種々のヘッダー情報が位置している。ヘッダー情報の終わりには“#” 1文字からなる行が位置しており、それ以降がラベル情報である。ラベル情報は

ラベルの時間位置（区切記号） フォント情報（区切記号） ラベル

の形式で記録されている。区切記号はスペースである。このうちフォント情報は Entropic 社のソフトウェア xwaves を利用した際にラベルを表示するフォントを指定する情報であり、分節音ラベルとしての意味はない。

分節音ファイルからラベル情報だけを抽出したければ、3フィールドからなるレコードで、第1フィールドが数字で構成されているものを取りだせばよい。

図1に分節音ファイルの例を示す。対象は模擬講演であり、その冒頭「テーマ、無人島に持っていくもの」という発話に該当する部分が示されている。「テーマ」と「無人島」の間に転記基本単位の境界がある。図2は、図1の発話中、「無人島に持っていくもの」に該当する区間の音声信号とサウンドスペクトログラムを分節音ラベルと時間的に同期させた形で示している。

詳しくは4.2以下で説明するが、分節音ラベルはすべて当該要素の終端時刻に付与されている。そのため、分節音の持続時間を知るには先行するラベルとの時間差を計算する必要がある。しかし、以下の説明からわかるように、ある分節音ラベルの直前のラベルが直前の分節音の終端を示しているとは限らないので（種々の補助ラベルであることがあるので）、持続時間を正確に計算するためには本文書で説明するラベル間の関係を理解しておくことが必要である。

分節音ファイルは xwaves のほかに、wavesurfer (<http://www.speech.kth.se/wavesurfer/>), Praat (<http://www.fon.hum.uva.nl/praat/>)等のソフトウェアでも利用できる。

（付記）CSJ 第二刷からは、Praat 用アノテーション形式である TextGrid 形式でもデータも提供しており、その中に分節音ラベルも格納されている。

signal S07M0833		← ヘッダー情報の始端
type 0		
comment created using xlabel Mon Mar 28 00:00:00 2003		
color 121		
font -*times-medium-r-*_*-17-*_*_*_*_*_*_*_*_*_*		
separator ;		
#		← ヘッダー情報の終端
0.315932	121 #,<cl>	← ラベル情報の始端, ラベルの融合
0.369405	121 t	
0.453561	121 eH	
0.557414	121 m	
0.699691	121 a	
1.121812	121 #	← 発話単位(転記基本単位)の始端
1.179131	121 m	
1.234524	121 u	
1.289000	121 zj	
1.382435	121 i	
1.446980	121 N	
1.480697	121 <cl>	
1.508634	121 t	
1.624114	121 oH	
1.669000	121 nj	
1.717077	121 i	
1.809000	121 m	
1.867724	121 o	
1.892772	121 <sv>	
2.013311	121 <cl>	
2.037185	121 Q,t	← ラベルの融合
2.059000	121 e	
2.120001	121 i	
2.142384	121 <cl>	
2.173719	121 k	
2.205756	121 u	
2.279000	121 m	
2.346767	121 o	
2.389000	121 n	
2.485180	121 o	

図1 分節音ファイルの構造

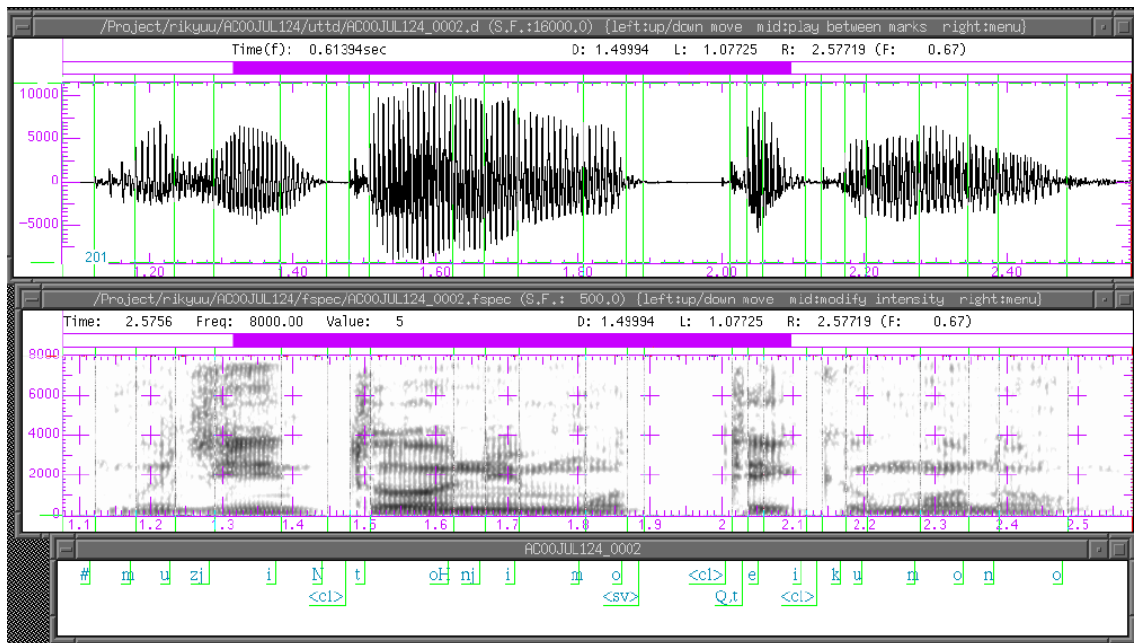


図2 ラベルと音声信号の関係

4. 2 ラベリングの基準

CSJ の分節音ラベリングでは、以下の基準にしたがってラベルの種類とラベル位置を決定している。しかし、境界の決定が不可能な場合はラベルの融合を許した。融合については 5.1 で詳しく説明する。

A) 発話単位に関するラベル

音声信号は転記テキストにおける転記基本単位によって発話単位に分割される（転記基本単位については transcription.pdf 参照）。分節音ラベリングでは発話単位の先頭に“#”を付与している（図 1 参照）。そのため、200ms 以上のポーズは分節音ラベリングの対象となっていない。発話単位中に（200ms 未満の）ポーズがあれば、その終端にラベル“<pz>”が付与されている。

B) 子音ラベルおよび関連する補助ラベル

破裂音：閉鎖区間の終端に補助ラベル“<cl>”を、後続する母音の始端に分節音ラベル“p”，“b”等を付与する（図 2 参照）。有声子音（特にガ行子音）が有声摩擦音ないし鼻音として実現されている場合も破裂音としてのラベリングを施す。この場合、“<cl>”は“<cl>,g”のように破裂音ラベルと融合させる（5.1 および 5.2 参照）。

摩擦音：摩擦ノイズの終端に分節音ラベル“s”，“h”，“z”等を付与する（図 4 参照）。

破擦音：閉鎖区間の終端に補助ラベル“<cl>”を、摩擦ノイズの終端にラベル“c”，“cj”，“z”，“zj”を付与する（図 4 参照）。“z”は、摩擦音のラベルと同一であるが、“<cl>”と共起することによって破擦音であることが示される（5.2 参照）。

鼻音：鼻音区間の終端に音素ラベル“m”，“n”等を付与する（図 2, 3, 4 参照）。

弾音：弾音に対応するスペクトル変化区間の終端にラベル“r”を付与する。ラ行子音が有声破裂音として実現されている場合も、弾音と同じラベル“r”を閉鎖区間の終端に付与する。

接近音：定常的なスペクトル区間をもたないことが接近音の特徴である。(1)先行母音のフォルマントの定常部終端ないし典型部（フォルマントがピークをなす時刻）、(2)接近音区間の中心部（フォルマントがピークをなす時刻）、(3)後続母音の定常部始端ないし典型部を決定し、(2)から(3)へのフォルマント遷移区間の中央にラベル“y”，“w”を付与する。また、先行母音の終端は(1)から(2)への遷移部の中央とする（図 3 参照）。また下の拗音についての説明も参照。

促音：後続する子音によってラベリングが変化する。後続子音が破裂音ないし破擦音の場合は、閉鎖区間の終端に補助ラベル“<cl>”を付与し、後続する母音の始端に“Q,t”のような促音ラベル“Q”と破裂音ラベル“p”，“t”等の融合ラベル（5.1 参照）を付与する（図 2 参照）。後続子音が摩擦音の場合は、摩擦ノイズの終端に“Q,s”のような促音と摩擦音の融合ラベルを付与する。ただし、「(F アッ)」「(D コッ)」など、転記タグ(F)(D)が付与された要素の末尾に出現する促音については、上記の規定に関わらず、“a,Q”“o,Q”のように促音とその前の母音を融合処理する（転記タグについては transcription.pdf 参照）。

撥音：後続する分節音によってラベリングが変化する。後続要素が鼻音であれば多くの場合に鼻音ラベルと融合処理する。ただし、稀に、撥音と後続鼻音との間に語境界が存在している場合などに、撥音と鼻音の間に音声信号の振幅差やフォルマントの異同が存在することによって、音声信号に明らかな不連続性が認められることがある。その場合は、撥音と鼻音を分離してラベリングする。後続子音が無声子音である場合は、有声区間の終端にラベル“N”を付与する。有声子音が後続する場合は、鼻音フォルマントの終端にラベルを付与する。母音が後続する場合で、アンチ・フォルマントなどによって鼻音の区間を決定できる場合は、そこにラベルを付与する。

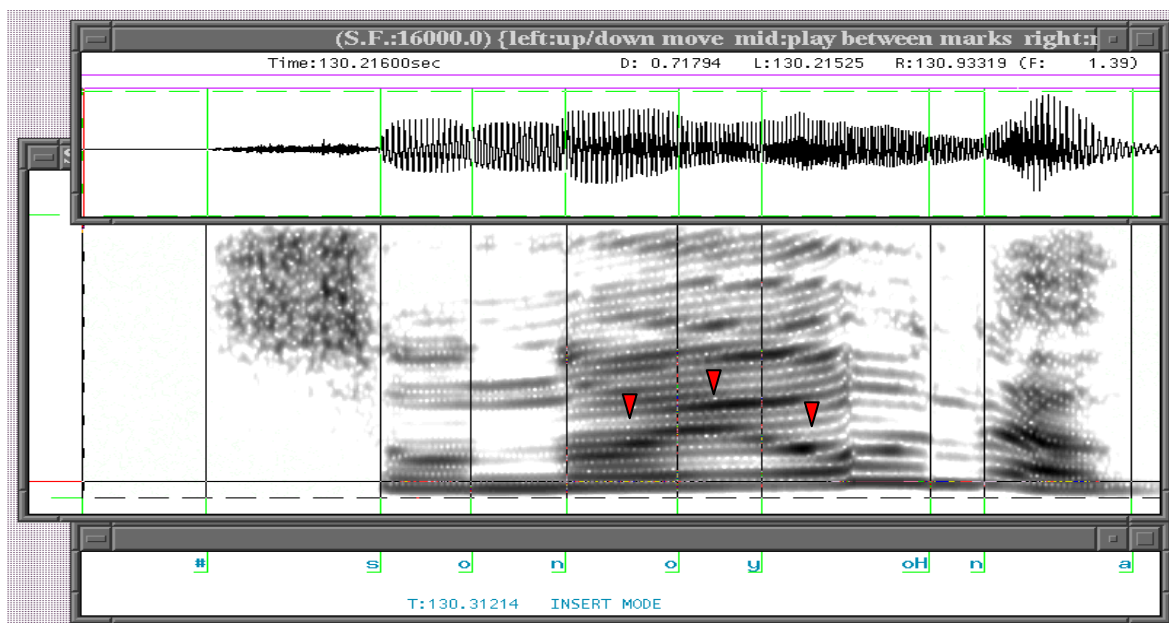


図3 接近音（ヤ行子音）“y”のラベリング例（女声による「そのような」）
 “o”と“oH”を結ぶフォルマント遷移区間の中央にラベル“y”が付与されている。
 図中の矢印は左から順に本文で説明した(1), (2), (3)の時刻を示している。

拗音：拗音の子音は、上に述べた直音の子音と同じ基準によって子音区間の終端を定め、そこに拗音ラベル“ky”，“cy”，“sy”，“my”等を付与する。図4に示されているように、子音区間から後続母音へのフォルマント遷移区間は拗音ラベルの区間に含まれていないことに注意。また、鼻音の拗音（“my”，“ny”）に関して、子音固有の区間を決定することは困難であるがフォルマントの遷移が認められれば、図3におけるラベリングと同様、遷移区間の中心にラベルを付与する。

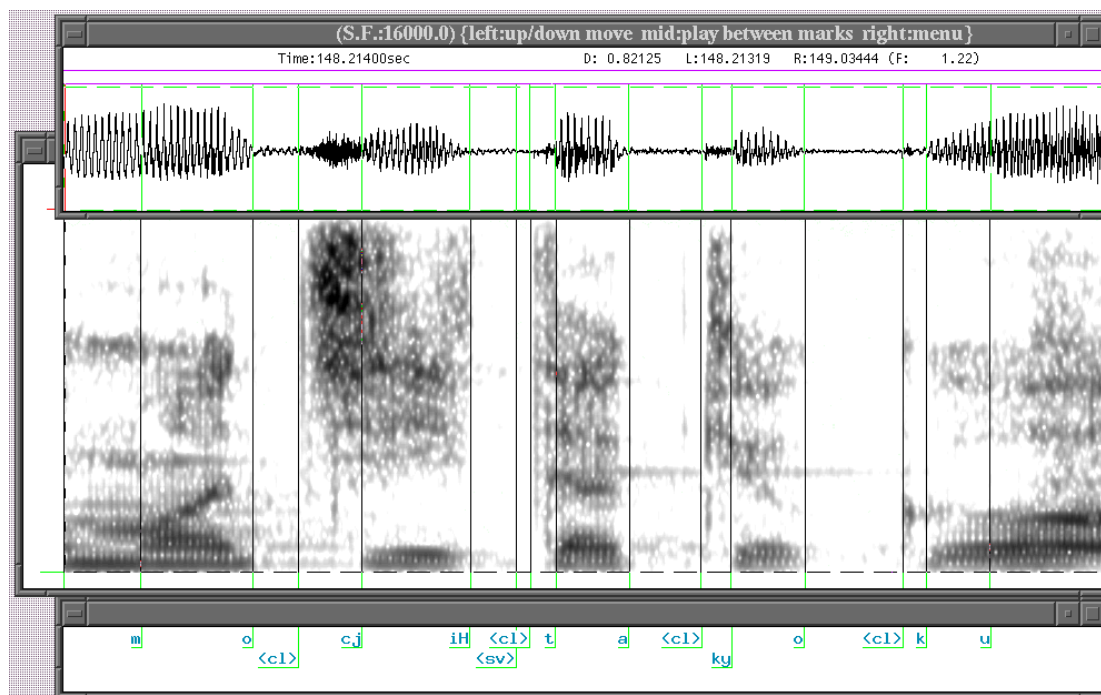


図4 拗音“ky”のラベリング例（女声による「用いた曲」）
 後続母音“o”へのフォルマント遷移区間はラベル“ky”の区間に含まれていない。

C) 母音ラベルおよび関連する補助ラベル

一般：母音にポーズや無声子音が後続している場合は、有声区間の終端に母音ラベル“a”，“i”等を付与する。

声帯振動の継続(“<sv>”)：母音に対応するフォルマントが消失した後に、スペクトルの低域に声帯振動だけが継続している場合は、フォルマントが認められる区間の終端に母音ラベルを与え、声帯振動の終端に補助ラベル“<sv>”を付与する（図2参照）。

共鳴の継続(“<uv>”)：上とは反対に声帯振動が停止した後にスペクトルの高域にフォルマントが継続して存在している場合は、声帯振動の終端に母音ラベルを与え、（無声の）フォルマントの終端に補助ラベル“<uv>”を付与する。

母音の連鎖：2個以上の母音がポーズを挟まずに連続して発音されている場合は、母音間のスペクトル遷移の中心に両母音の境界を認め、先行母音のラベルを付与する。

ボーカル・フライ：母音の終わりに非周期的な声帯振動、いわゆるボーカル・フライが生じることがある。その場合周期的な声帯振動区間の終端に母音ラベルを付与し、それに後続するボーカル・フライ区間の終端に補助ラベル“<fr>”を付与する。なお、分節音ラベル“<fr>”と転記テキストにおける<FV>タグの生起とは必ずしも一致しない。

D) その他の補助ラベル

雑音(“<N>”および“”)：発話と重ならず種々の雑音が単独で生じている場合、その終端に付与する。話者の呼吸音についてはラベル“”を、その他の雑音にはラベル“<N>”を利用する。雑音については転記テキスト中にも所在が記録されている (transcription.pdf 参照)が、分節音ラベリングにおける雑音と一致するとは限らない。なお、学会講演のなかには、空調設備や工事の騒音などが音声ファイルの全体にわたって音声と重畳しているものがある。この種の雑音には、記録表データのなかで主観評価を与えている (data_attribute.pdf 参照)。

5. 利用上の注意

5. 1 ラベルの融合

前節に述べた基準では分節音の境界を確定できないことがある。その場合、無理にラベル位置を確定することはせずに、存在すると考えられる複数のラベルをひとつのラベルにまとめて、位置を確定できる最初の（ラベル列としては最後の）分節音の時刻に付与する。このようなラベルを融合ラベルと呼び、融合ラベルを付与する作業を融合処理と呼ぶ。本節では、融合処理が生じ易い環境について解説する。

融合処理が頻繁に生じる環境のひとつは母音が無声化した場合である。「シ」「ス」のような無声摩擦音を含むモーラに母音の無声化が生じると、モーラの全体が無声摩擦子音として実現されるが、そのうちどこまでが子音音素で、どこからが無声化母音であるかを時間軸上で区分することはできない。この場合、母音が無声化したモーラの終端に、融合ラベル“sj,I”ないし“s,U”を付与している。融合ラベルでは、融合したラベルを、カンマで区切って想定される出現順に並べ、ひとつのラベルとしている。

ラベルの融合が生じやすい環境には母音の無声化のほかに、二重母音を含めた母音連鎖、促音、撥音、半子音(“w”, “y”)、補助記号“<cl>”などがある。母音連鎖では、“a,u”, “o,a”のような融合ラベルが生じる。促音(“Q”)は、4.2でも述べたように多くが後続ラベルと融合し、“Q,t”, “Q,s”のよ

うな融合ラベルとなる。図 1, 2 では 2.03 秒付近に “Q,t” が生じている。撥音(“N”)では、鼻子音が後続する場合に融合が頻繁で、“N,m” , “N,n” のようなラベルが生じる。補助ラベルの “<cl>” は融合することの非常に多いラベルであるが、これについては 5.2 で別途説明する。

ラベルの融合は 3 個以上のラベルにわたって生じることがある。例えば“Q,k,U”は促音に後続するモーラ「ク」の母音が無声化している場合、“Q,k,U,s,U”は促音に後続する「クス」の母音がともに無声化したうえ、さらに子音 “k” , “s” 間の境界も確定できない場合である。

上記のほかに、“h” , “r” , “n” , “m” の前後、発話末で音声信号の振幅が小さい区間、笑いながら発話している区間、ささやき声での発話区間などでも融合処理が多くなりやすい。

5. 2 <cl>について

すでに説明したように、“<cl>” ラベルは破裂音ないし破擦音における閉鎖区間の終端時刻に付与される補助ラベルであり、分節音を構成する音声要素のラベルである。本節では、本ラベルと “p” , “b” , “t” , “d” , “c” , “z” 等の子音ラベルとの関係について説明する。

無声破裂音 (“p” , “t” , “k” およびその口蓋化子音と拗音子音) を含む CV モーラでは、閉鎖の開放後に氣息音(aspiration)が続き、その後で、母音の声帯振動が始まるのが典型的なパターンである。この場合、“<cl>” は閉鎖の終端に生じるバースト音 (スパイク) の始端に付与され、子音ラベルは声帯振動 (後続母音) の開始時点に付与される。したがって、“<cl>” と子音ラベルの時間間隔が氣息音の持続時間 (いわゆる VOT) に該当する。しかし、“<cl>” と子音ラベルが常に時間軸上で分離できるとは限らない。つまり無声破裂音であっても VOT が実際上ゼロになることがある。その場合、“<cl>” と子音ラベルは融合ラベルとなって同一の時刻 (声帯振動の開始時点) に付与されている。

有声破裂音 (“b” , “d” , “g” およびその口蓋化子音と拗音子音) を含むモーラでは、閉鎖の開放と同時にもしくはそれよりも早い時刻に声帯振動が始まるのが普通である。その場合、“<cl>” は子音ラベルと融合して、バースト音の時刻に付与される。有声破裂音の発音では閉鎖区間の途中で声帯振動が始まることもあるが、閉鎖区間中における声帯振動の開始点にはラベルを付与していない (そのため、いわゆる負の VOT は計測できない)。

無声破擦音 (“c” およびその口蓋化子音と拗音子音) では、“<cl>” は破擦音の前半を構成する閉鎖区間の終端に付与される。

有声破擦音 (IPA の [dʒ] と [dʒ]) は有声摩擦音 (IPA の [z] と [ʒ]) とともにザ行子音の異音をなす。CSJ の分節音ラベルでは、ザ行子音 (の終端) にはラベル “z” が付与されており、そのうち有声破擦音であるものには、閉鎖の終端に “<cl>” が付与されている。

その他 “<cl>” について注意が必要な場合を列挙しておく。転記基本単位の冒頭に破裂音が位置している場合、それが無声閉鎖音であれば融合ラベル “#,<cl>” が用いられている (ただし吸気音の存在等によって閉鎖区間の始端が推定できる場合は融合処理を施していない)。子音が有声閉鎖音で、閉鎖の開放に先立って prevoicing が明瞭に観察される場合は、その開始時刻に “<cl>” を付与している。“<pz>” に破裂音が後続する場合も同様の処理を施している。語中や語境界に “<sv>” や “<uv>” が生じ、これらに破裂音が後続する場合、融合処理されていることが多い。

無声破裂音、特に軟口蓋破裂音では、バースト音が 2 回以上生じることがある。その場合、“<cl>” は振幅の大きいバースト音 (振幅に差がなければ最後のバースト音) に付与している。

5. 3 転記テキスト中の<H>, <Q>について

CSJ の転記テキストの発音形では<H>,<Q>というタグが用いられており、これらはそれぞれ「非語彙的な母音の延長」と「非語彙的な子音の延長」を表している。例えば「私」「本当」が「ワターシ」「ホンット」と発音されていれば、転記テキストは「ワタ<H>シ」「ホン<Q>ト」となる（詳細は transcription.pdf 参照）。

しかし、分節音ラベリングでは、これらの音声学的な長母音、長子音（促音）には長母音ラベル“H”ないし促音ラベル“Q”を付与していない。分節音ラベルにおいて“H”，“Q”が付与されているのは語彙的に長母音ないし促音という音素（すなわち<>に囲まれていない音素“H”，“Q”）が指定されている場合だけである。発音形音素表記における/a<H>/と/<Q>to/は、分節音ラベルとしては、短母音“a”、短子音“t”として処理されている。転記テキストにおいて<H>ないし<Q>が付与された分節音だけを選択的に調査したい場合は、転記テキストを参照する必要がある。このような複合的な情報検索の便宜を図るために、CSJ では各種付加情報を統合した XML 文書を提供している (xml.pdf 参照)。

6. ラベリング作業の実施方法

CSJ の分節音ラベリング作業は、以下の手順で実施した。1)から 5)までには複数の作業者が介在している。作業の詳細については以下に示した参考文献参照。

- 1) 転記テキストの「発音形」情報から発話に含まれる音素列を自動生成する。
- 2) 隠れマルコフモデルを用いて自動的に上記音素列と音声信号のアラインメントをとり、これを分節音ラベルの初期値とする。
- 3) ラベリング一次作業者が上記初期値を手で修正して本文書に解説されているラベル体系に変換する。
- 4) 手で修正されたラベルを二次作業（音声学の専門家 2 名）が検討して修正する。
- 5) 上記 4)において、転記テキストに問題が発見されれば、それを修正する。

7. Phone について

5.1 に述べたように、CSJ の分節音ラベリング体系では、分節音の境界を確定できない場合に融合ラベルを認めている。しかしながら、処理の都合上、時刻情報の信頼性は低くてもよいから単独のラベル（の連続）として扱いたい、というケースがしばしば生じる。そこで、そのような単位として Phone を設定した。Phone は、分節音ラベルからおおよそ以下の手順に従って生成される（以下、“|”で分節音および Phone 境界を表す。また、着目している箇所には下線を施す）。

- 1) “<cl>|Q,t” または “<cl>,Q,t” とラベリングされているものについて、“Q”を“<cl>”の前に移動させ両者を融合処理する。

例: |<cl>|k,A|<cl>|Q,t|oH| → |<cl>|k,A|Q,<cl>|t|oH|

- 2) 融合ラベルを分割する。境界の時刻情報は、当該時間区間を等分割して得られたものとする。

例: |<cl>|k,A|Q,<cl>|t|oH| → |<cl>|k|A|Q|<cl>|t|oH|

- 3) 長母音を分割する。この場合も等分割する。

例: |<cl>|k|A|Q|<cl>|t|oH| → |<cl>|k|A|Q|<cl>|t|o|H|

Phone は waves 形式のファイルとして提供されるわけではないが（提供されるのは本文書 6 節までに示した方法で認定した分節音情報を納めた分節音ファイル.seg だけである）、研究用付加情報のなかには Phone という概念を利用するものがある。そのうちの一つはイントネーションラベルである（intonation.pdf 参照）。図 5 に示すように、単語ラベル（イントネーションラベルの一つ）の終端は、それを構成する最後の Phone の終端に一致させている。

単語	noHeN				nji		iQ		ta			
分節音	n	oH		e	N,nj		i,i	<cl>	Q,t	a		
Cf. Phone	n	o	H	e	N	nj	i	i	Q	<cl>	t	a

図 5. 単語・分節音ラベリング例（模式図）「農園に行った」

このように単語境界に融合ラベルが生じる場合は、単語境界と分節音境界が一致しない点に注意する必要がある。

また、各種研究用付加情報を統合した XML 文書では、分節音に関する情報を Phone 情報として記述している。融合処理された分節音ラベルは直接的には記述していないが、Phone 関連の情報から復元することが可能である（詳細は xml.pdf 参照）。

8. XML 文書および音響モデルにおける「音素」について

分節音ラベリングの過程では、音素(phoneme)という概念は、自動アラインメントの過程で潜在的に利用されるだけである。しかし分節音ラベル以外の研究用付加情報のなかには音素という概念を利用するものがある。それは、音声認識用音響モデルと XML 文書である。

このうち音響モデルにおける音素セットについては asr.pdf に解説がある。XML 文書では階層構造の一部として Phoneme 要素が設定されており、その子要素として Phone 要素が配置されている（詳細は xml.pdf 参照）。この Phoneme 要素は、自動アラインメント作業のために転記テキストから自動生成した記号列のことであり、分節音ラベルから補助ラベル類を除外したものにほぼ該当する。音韻論的に厳密な意味での phoneme ではない。

参考文献

菊池英明・前川喜久雄・五十嵐陽介・米山聖子・藤本雅子「日本語話し言葉コーパスの音声ラベリング」音声研究, 7 (3), pp.16-26, 2003.

付録 現代日本語のモーラと CSJ 分節音ラベルの対応

直音系列モーラ	拗音系列モーラ	周辺のモーラ
アイウエオ a i u e o	ヤユヨ ya yu yo	イエ ye
カキクケコ ka kji ku ke ko	キャ キュ キョ kya kyu kyo	クワ kwa
ガギグゲゴ ga gji gu ge go Ga Gji Gu Ge Go	ギャ ギュ ギョ gya gyu gyo Gya Gyu Gyo	グワ gwa Gwa
が き ぐ げ ご @a @ji @u @e @o	ギャ ギュ ギョ @ya @yu @yo	(注：有声破裂音) (注：有声摩擦音)
サシスセソ sa sji su se so	シャ シュ ショ sya syu syo	シェ スイ sye si
ザジズゼゾ za zji zu ze zo	ジャ ジュ ジョ zya zyu zyo	ジェ ズイ zye zi
タチツテト ta cji cu te to	チャ チュ チョ cya cyu cyo	テイ トウ チェ ツア ツイ ツエ ツオ テュ ti tu cye ca ci ce co tyu
ダデド da de do		デイ ドウ デュ di du dyu
ナニヌネノ na nji nu ne no	ニャ ニュ ニョ nya nyu nyo	ニエ nye
ハヒフヘホ ha hji fu he ho	ヒャ ヒュ ヒョ hya hyu hyo	ヒェ ファ フィ フェ フォ フュ hye fa fi fe fo fyu
バビブベボ ba bi bu be bo	ビャ ビュ ビョ bya byu byo	ヴァ ヴィ ヴ ヴェ ヴォ va vi vu ve vo
パピプペポ pa pi pu pe po	ピャ ピュ ピョ pya pyu pyo	
マミムメモ ma mi mu me mo	ミャ ミュ ミョ mya myu myo	ミエ mye
ラリルレロ ra ri ru re ro	リャ リュ リョ rya ryu ryo	
ワ ヲ wa o	ウィ ウェ ウォ wi we wo	

撥音 促音 長母音 (第二要素)

N Q H

(注) この表は現代日本語に生じうるモーラを示したものであり、そのすべてが CSJ に生じているとはかぎらない。