

韻律研究のための日本語話し言葉コーパス XML 文書作成  
Construction of XML document for the study of prosody  
using *the Corpus of Spontaneous Japanese*

菊池 英明<sup>†</sup>      前川 喜久雄<sup>‡</sup>  
KIKUCHI Hideaki<sup>†</sup>      MAEKAWA Kikuo<sup>‡</sup>

<sup>†</sup> 早稲田大学人間科学学術院      <sup>‡</sup> 国立国語研究所研究開発部門

<sup>†</sup> Faculty of Human Sciences, Waseda University

<sup>‡</sup> Dept. Language Research, National Institute for Japanese Language

kikuchi@waseda.jp

**Abstract:** Based upon the XML document of the Corpus of Spontaneous Japanese, we devised new XML document that expresses explicitly the prosodic structure of utterances. The half of this paper is devoted for the discussion of design and implementation issues of the new document, which is followed by the presentation of the result of pilot analysis regarding the linguistic variation of some boundary pitch movements in Japanese.

## 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、話者情報、転記情報、形態論情報、分節音情報、韻律情報、係り受け情報など様々な研究用の情報が収められ、さらにこれらの情報を相互に関係付けて XML 形式の文書としたものが公開されている[1]。同一の音声データに対して付加された種類の異なる研究用付加情報を、相互の依存関係を表現した形で記述することにより、例えばある語の統語的な環境によるアクセント実現位置の分布の違いを調べるなど、種類の異なる情報に跨った検索が容易になる。

一方で、日本語のイントネーションにおいてはアクセント句が重要な単位となって主に言語的情報を伝達しており、したがってアクセント句単位を仮定した音声学的分析の意義は大きい。既に公開されている CSJ の XML 文書においては、広く知られている日本語イントネーション記述モデルの J\_ToBI を拡張した X-JToBI[2]によって記述された韻律的境界強度(Break Index、以下 BI)の情報を用いれば簡易的にアクセント句単位を仮定した分析を行うことが可能である。しかしながら、例えばアクセント句

境界やアクセント核などの種々のトーンイベントを記述する韻律ラベル情報は、時間的に包含される分節単位情報に関係付けて記述されているため、一部のトーンイベントの情報が音韻論的に分節単位情報と関係付けられず、その結果として例えばアクセント句の始端の正確な時間位置情報を形態論情報や分節単位情報と関係付けて得ることが困難になっている。そこで、我々は韻律ラベル情報を時間的対応関係だけでなく音韻論的対応関係を反映させる形で XML 文書内に表現して、より幅広い研究用途に供することにした。

以下には、2 章に CSJ の XML 文書の基本的な仕様を解説し、3 章にアクセント句単位に基づいて再構成した XML 文書の仕様を述べる。4 章で今回再構成した XML 文書を用いた予備的な分析を通じて、提案する仕様の有効性を検討する。

## 2. 日本語話し言葉コーパスにおける XML 文書

CSJ のコアと呼ばれる一部のデータには、話者情報、転記情報、文節情報、形態論情報、分節音情報、韻律情報、節単位情報、係り受け情報、談話境界情報、重要文情報といった多様な種類の研究用付加情

報が含まれている[3]。これらの情報には、言語学的情報と音声学的情報が混在しており、例えば語の境界を挟んで音声的に融合しているために分節単位が語境界をまたぐような場合など、階層構造を逸脱するものもある。しかしながら、整合性検証や検索などの多くの用途においては言語学的な階層構造を想定することが予想されるため、CSJの研究用情報を記述するには階層構造を基本とすることにした。公開されているXML文書における各要素の関係を図1に、各要素の内容の一部を表1に、XML文書の例を図2に示す。

XMLに記述されたデータの構造を変換するためのXML関連規格XSLT(Extensible Style Language Transformation)を用いれば、こうした階層構造で表現された種々の情報から必要な情報だけを取り出すことが容易になる。図3に、「形容詞の活用形とアクセント位置の情報を抽出する」XSLTを示す。ここで、CSJのXML文書において語(CSJでの短単位)はSUW要素、その品詞はSUWPOS属性、代表表記はSUWLemma属性、活用形はSUWConjugateForm属性、発音はPhoneticTranscription属性に記録される。また、語の知覚アクセント位置は韻律情報から得られ、

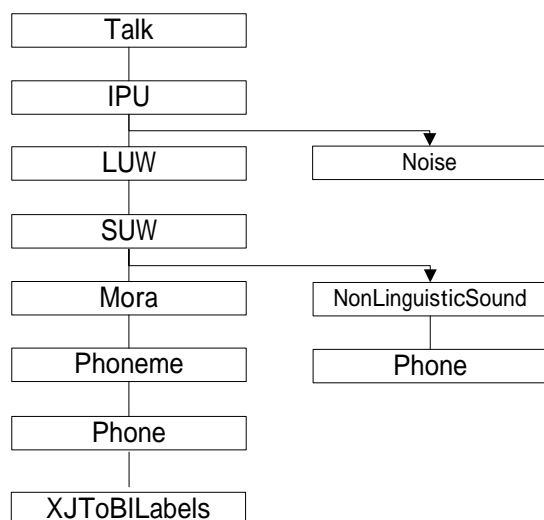


図 1: XML 要素関連図

XJToBILabelWord 要素の PerceivedAccPos 属性に記録されている。図で示した XSLT には、まず文書中から SUW 要素を探し出し、SUWPOS 属性の値が「形容詞」である場合に、SUWLemma, SUWConjugateForm, PhoneticTranscription 属性の値をカンマ区切りで出力し、さらに下位層にある XJToBILabelWord 要素の PerceivedAccPos 属性の値を出力するという手続きが記述されている。

表 1: XML 要素の内容

要素	属性	説明
Talk		講演
	RecordingDate	録音日
	SpeakerID	話者ID
	BirthDate	誕生日
	WaveFilePath	音声ファイルパス
IPU		転記基本単位
	Channel	音声録音チャンネル
	IPUStartTime	開始時刻
	IPUEndTime	終了時刻
LUW		長単位
	LUWPOS	品詞(長単位)
	LUWConjugateType	活用型(長単位)
	LUWConjugateForm	活用形(長単位)
	LUWDictionaryForm	代表形(長単位)
	LUWLemma	代表表記(長単位)
SUW		短単位
	SUWDictionaryForm	代表形(短単位)
	SUWLemma	代表表記(短単位)
	SUWPhoneTrans	発音形(短単位)
	SUWPOS	品詞(短単位)
	SUWConjugateType	活用の種類(短単位)
	SUWConjugateForm	活用形(短単位)
	LexicalAccPos	語彙アクセント位置
	TagDisfluency	語断片 (D)
	TagFiller	フィラー (F)
	TagIncorrect	言い誤り (W)
	TagIncorrectNorm	言い誤り正式発音
	APID	所属アクセント句の ID
Mora		
	MoraEntity	モーラ記号
	Uncertain	発音不明瞭 (?)
	PerceivedAccPos	知覚アクセント位置
Phoneme		音素
	PhonemeEntity	音素記号
Phone		分節音
	PhoneEntity	分節音記号
	Devoiced	無声化の有無
	PhoneStartTime	開始時刻
	PhoneEndTime	終了時刻
	StartPosUncertain	開始位置不明
	EndPosUncertain	終了位置不明
XJToBILabels		韻律ラベル
	Entity	ラベル文字列
	Time	ラベル付与位置(時間)
Noise		言語音と独立した非言語音
NonLinguisticSound		言語音と共起する非言語音



```

<Talk RecordingDate="2000-01-01" SpeakerID="0001"
BirthDate="1950-01-01" WaveFilePath="wav/S03f0119.wav">
  <IPU Channel="L" IPUStartTime="244.050" IPUEndTime="245.009">
    <LUW LUWPOS="代名詞" LUWDictionaryForm="イツ"
LUWLemma="何形">
      <SUW SUWPOS="代名詞" SUWDictionaryForm="イツ"
SUWLemma="何形" LexicalAccPos="I" APID="304">
        <Mora MoraEntity="イ" PerceivedAccPos="I">
          <Phoneme PhonemeEntity="i">
            <Phone PhoneEntity="i" PhoneStartTime="244.073871"
PhoneEndTime="244.154540">
              <XJToBILabelTone Time="244.092331" F0="210.791">%L
<XJToBILabelTone>
              <XJToBILabelTone Time="244.114585" F0="216.325">A
<XJToBILabelTone>
            </Phone>
          </Phoneme>
        </Mora>
        <Mora MoraEntity="ツ">
          <Phoneme PhonemeEntity="c">
            <Phone PhoneEntity="cl" PhoneStartTime="244.154540"
PhoneEndTime="244.187874">
              <Phone PhoneEntity="c" PhoneStartTime="244.187874"
PhoneEndTime="244.240331">
            </Phoneme>
          <Phoneme PhonemeEntity="u">
            <Phone PhoneEntity="u" PhoneStartTime="244.2440331"
PhoneEndTime="244.268501">
            </Phoneme>
          </Mora>
        </SUW>
      </LUW>
    <LUW LUWPOS="助詞" LUWDictionaryForm="モ" LUWLemma="
も">
      <SUW SUWPOS="助詞" SUWDictionaryForm="モ" SUWLemma="
も" APID="304">
        <Mora MoraEntity="モ">
          <Phoneme PhonemeEntity="m">
            <Phone PhoneEntity="m" PhoneStartTime="244.268501"
PhoneEndTime="244.328683">
            </Phoneme>
          <Phoneme PhonemeEntity="o">
            <Phone PhoneEntity="o" PhoneStartTime="244.328683"
PhoneEndTime="244.372218">
            </Phoneme>
          </Mora>
        </SUW>
      </LUW>
    <LUW LUWPOS="助詞" LUWDictionaryForm="ノ" LUWLemma="
の">
      <SUW SUWPOS="助詞" SUWDictionaryForm="ノ" SUWLemma="
の" APID="304">
        <Mora MoraEntity="ノ">
          <Phoneme PhonemeEntity="n">
            <Phone PhoneEntity="n" PhoneStartTime="244.372218"
PhoneEndTime="244.418315">
            </Phoneme>
          <Phoneme PhonemeEntity="o">
            <Phone PhoneEntity="o" PhoneStartTime="244.418315"
PhoneEndTime="244.494613">
              <XJToBILabelTone Time="244.494613" F0="151.82"
ToneClass="FBT">L% <XJToBILabelTone>
            </Phone>
          </Phoneme>
        </Mora>
      </SUW>
    </LUW>
  </IPU>
</Talk>

```

図2: XML化したCSJの研究用情報の例

```

<?xml version="1.0" encoding="EUC-JP"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xml:lang="ja">

<xsl:output method="text" indent="yes" encoding=
"EUC-JP"/>

<!--まず短単位を探す-->
<xsl:template match="SUW">
  <!--品詞が形容詞であるものを探す-->
  <xsl:if test="@SUWPOS='形容詞'">
    <xsl:value-of select="@SUWLemma"/>
    <xsl:text>,</xsl:text>
    <xsl:value-of select="@SUWConjugateForm"/>
    <xsl:text>,</xsl:text>
    <xsl:value-of select="@PhoneticTranscription"/>
    <xsl:value-of select="descendant::XJToBILabelWord[1]/
@PerceivedAccPos"/>
    <xsl:text>,</xsl:text>
    <xsl:text>#&#x0a;</xsl:text>
  </xsl:if>
</xsl:template>

</xsl:stylesheet>

```

図3: CSJのXML文書から「形容詞の活用形とアクセント位置の情報を抽出する」XSLTの例

一方で、例えばアクセント句境界やアクセント核などの種々のトーンイベントを記述する韻律ラベル情報は、時間的に包含される分節単位情報の下層に記述されているため、一部のトーンイベントの情報が音韻論的に分節単位情報と関係付けられず、その結果として例えばアクセント句の始端の正確な時間位置情報を形態論情報や分節単位情報と関係付けて得ることなどが困難になっている。例えばアクセント句が発話中で接続する場合に境界に位置するトーンイベントが形態論的な境界とずれるケースは、全アクセント句中20%近い割合で存在する。

そこで我々は、公開されているXML文書において基本となっている言語学的な階層構造をアクセント句単位を基本とする構造に変換することで、韻律研究用途に適したXML文書を作成した。次章には、新たなXML文書作成の手続きとそれによって作成されたXML文書の仕様を説明する。

### 3. アクセント句単位に基づく再構成

以下に、公開版XML文書を、アクセント句単位を基本とする構造に変換する再構成の手続きを示す。

なお、再構成後の XML 文書における要素間の関係は図 4 のようになる。

### (1) LUW 要素と IPU 要素の情報の移行

図 4 に示したとおり、Talk 要素と SUW 要素の間にアクセント句単位に相当する AP 要素を新規に構成し、代わりに LUW 要素と IPU 要素を取り除く。これは、AP 要素の境界と LUW, IPU 各要素の境界が一致しないことが多いためである。LUW, IPU 要素の属性は SUW 要素に移行させる。

### (2) アクセント句単位の構成

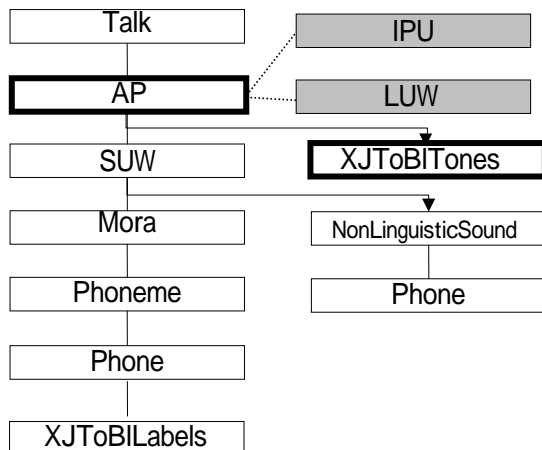
図 2 に示したように、公開版 XML には、既にアクセント句単位に相当する AP 要素の ID が、SUW 要素の APID 属性として記述されている。この値が同一である SUW 要素どうしを AP 要素の下位にまとめあげていく。なお、APID 属性の値は韻律ラベル情報のうちの BI 層のラベルが“2”以上の値をとる際にインクリメントして与えられている。

### (3) トーンラベル情報の音韻論的対応付け

前章で述べたように、分節音単位情報との関係において音韻論的な対応と時間的な対応が異なる場合がある。このような場合も含めて、アクセント句単位に相当する AP 要素の下位に、音韻論的に対応するトーンラベル情報の複製を位置づける。

具体的には、分節単位情報に記述されている時間情報から音韻論的に定まるアクセント句境界の時間位置を取得し、その近傍にあるトーンラベル情報について、想定されるトーンラベル情報の文法に従うようにアクセント句単位に関係付けていく。

以上の手続きを経て再構成した XML 文書の例を図 4 に示す。このようにして再構成した XML 文書を用いれば、各アクセント句がどのような境界音調で終わっているかを形態論情報などと関係付けて分析することが容易になる。



```
<Talk RecordingDate="2000-01-01" SpeakerID="0001"
  BirthDate="1950-01-01" WaveFilePath="wav/S03f0119.wav">
  <AP APID="304">
    <XJToBITone Time="244.098746" F0="210.791"
      ToneClass="IBT">%L</XJToBITone>
    <XJToBITone Time="244.114585" F0="216.325">A </XJToBITone>
    <XJToBITone Time="244.485417" F0="151.82"
      ToneClass="FBT">L%</XJToBITone>
    <SUW SUWPOS="代名詞" SUWDictionaryForm="イツ"
      SUWLemma="何時時" LexicalAccPos="1" APID="271" Channel="L"
      IPUStartTime="244.050" IPUEndTime="245.009" LUWPOS="代名
      詞" LUWDictionaryForm="イツ" LUWLemma="何時時">
    <Mora MoraEntity="イ" PerceivedAccPos="1">
    <Phoneme PhonemeEntity="i">
    <Phone PhoneEntity="i" PhoneStartTime="244.073871"
      PhoneEndTime="244.154540">
    <XJToBILabelTone Time="244.092331" F0="210.791">%L
    </XJToBILabelTone>
    <XJToBILabelTone Time="244.114585" F0="216.325">A
    </XJToBILabelTone>
    </Phone>
    </Phoneme>
  </Mora>
  <Mora MoraEntity="ツ">
    <Phoneme PhonemeEntity="c">
    <Phone PhoneEntity="cl" PhoneStartTime="244.154540"
      PhoneEndTime="244.187874">
    <Phone PhoneEntity="c" PhoneStartTime="244.187874"
      PhoneEndTime="244.240331">
    </Phoneme>
    <Phoneme PhonemeEntity="u">
    <Phone PhoneEntity="u" PhoneStartTime="244.244.240331"
      PhoneEndTime="244.268501">
    </Phoneme>
  </Mora>
  </SUW>
  <SUW SUWPOS="助信詞" SUWDictionaryForm="ㇿ" SUWLemma="
  ㇿ" APID="304" Channel="L" IPUStartTime="244.050"
  IPUEndTime="245.009" LUWPOS="助信詞"
  LUWDictionaryForm="ㇿ" LUWLemma="ㇿ">
  <Mora MoraEntity="ㇿ">
  <Phoneme PhonemeEntity="m">
  <Phone PhoneEntity="m" PhoneStartTime="244.268501"
  PhoneEndTime="244.328683">
  </Phoneme>
  <Phoneme PhonemeEntity="o">
  <Phone PhoneEntity="o" PhoneStartTime="244.328683"
  PhoneEndTime="244.372218">
  </Phoneme>
  </Mora>
  </SUW>
  <SUW SUWPOS="助信詞" SUWDictionaryForm="ノ" SUWLemma="
  ㇽ" APID="304" Channel="L" IPUStartTime="244.050"
  IPUEndTime="245.009" LUWPOS="助信詞"
  LUWDictionaryForm="ノ" LUWLemma="ㇽ">
  <Mora MoraEntity="ノ">
  <Phoneme PhonemeEntity="n">
  <Phone PhoneEntity="n" PhoneStartTime="244.372218"
  PhoneEndTime="244.418315">
  </Phoneme>
  <Phoneme PhonemeEntity="o">
  <Phone PhoneEntity="o" PhoneStartTime="244.418315"
  PhoneEndTime="244.943113">
  <XJToBILabelTone Time="244.494613" F0="151.82"
  ToneClass="FBT">L%</XJToBILabelTone>
  </Phone>
  </Phoneme>
  </Mora>
  </SUW>
  </AP>
</Talk>
```

図 5: 再構成した XML 文書の例

#### 4. 再構成された XML 文書を用いた予備的分析

##### 4.1 アクセント句単位についての基本的な分析

本節では、前章で述べた再構成後の XML 文書について、アクセント句単位を中心にした基本的な分析の結果を示す。なお本節では CSJ のコアとよばれる講演のうち 174 講演の XML 文書を分析の対象とする。

まず、講演種別ごとのアクセント句単位数、平均時間長を調べたところ、表 2 に示すようにアクセント句単位数は 10 万を超えることがわかった。なお、分析対象としたデータにおいては、平均的には一転記基本単位あたり 2.43 アクセント句、一アクセント句あたり 3.16 短単位という構成であった。

また、平均時間長は 700ms 前後であり、講演種別ごとの平均時間長には有意な差が認められた。

なお、アクセント句が転記基本単位境界に跨るケースが 6781 事例、長単位がアクセント句境界に跨るケースが 9741 事例見られた。これらのことから、本稿で述べているアクセント句単位に再構成した XML 文書の利便性の一つが示唆されたといえる。

最後に、[2] で少量データを対象に調べた BPM (Boundary Pitch Movement) の出現頻度分布を、再構成した文書に対して調べたところ表 3 のような結果となった。[2] の結果に対して学会講演における "L%+H%" の割合がやや多くなっている以外は概ね傾向に変化がないことが確かめられた。

表 2 講演種別ごとのアクセント句単位分析結果

講演種別	講演数	アクセント句単位数	アクセント句時間長([ms])
学会講演	62	51109	675.3
模擬講演	106	74246	687.7
再朗読	6	6358	745.3
全体	174	131713	685.7

表 3 BPM の出現頻度分布

(( )内は各講演種別における BPM 総数に対する割合[%])

講演種別	L%+H%	L%+HL%	L%+LH%
学会講演	12926(85.1)	2152(14.1)	95(0.8)
模擬講演	9706(58.7)	6572(39.7)	237(1.6)
再朗読	1453(97.9)	30(2.0)	1(0.0)
全体	24085(72.6)	8754(26.3)	333(1.1)

##### 4.2 アクセント句末境界音調の変異

本節では再構成された XML 文書の本語韻律研究における有用性を示すために、句末境界音調における言語変異現象の分析例を示すことにする。前節に示したように東京語のアクセント句末には様々な境界音調が生じるが、そのうち上昇調(L%+H%)と上昇下降調(L%+HL%)には川上藁が「浮き上がり調」と名付けた変異形が生じる[4]。

通常の上昇調、上昇下降調におけるピッチの変動がアクセント句最終モーラ(ないし音節)の内部に収まっているのに対して、浮き上がり調では上昇の開始点がアクセント句末から数えて二つ目(penult)もしくは三つ目(anti-penult)のモーラにずれているのが特徴である。X-JToBI では浮き上がり調の境界音調を有するアクセント句に対しては、プロミネンス層に「FR」(floating rise)というラベルを付与する。

CSJ のコアには 22091 件の上昇調が生じている(注釈層に種々の理由でラベルの信頼性が低いことを示す AYOR ラベルが付与されたものを除外した件数である)。そのうち FR が付与されているものが 1539 件であるから、上昇調における浮き上がり調の生起率は約 7% になる。

これまで検討した範囲では、浮き上がり調の生起に影響しているのは音韻的要因よりもむしろアクセント句末の形態論的ないし統語的特徴である。例えばアクセント句末の 2 語(短単位)を分類基準として、どのような句に FR が多く付与されているかを集計すると、上位には「です+が」132 件、「けれど+も」121 件、「です+ね」116 件、「です+けど」61 件、「けど+も」40 件などが並ぶ。これら 5 例中「です+ね」を除く 4 例までがいわゆる逆接表現になっていることが注目される。

表 4 は FR 付与数の多い句末表現上位 10 句に対して、FR の付与率 (FR 付与数/上昇調生起数\*100) を計算した結果である。表中の句はいずれも平均よりは FR が生じやすいものであるが、句末が逆接表現の場合には浮き上がり調の生起率が一層高くなっていることがわかる。

表4 上昇調境界音調(L%H%)における浮き上がり調生起率とアクセント句末2語の関係

句末の2語	FR	L%H%	%FR	逆接
です+が	132	553	23.9	✗
けれど+も	121	436	27.8	✗
です+ね	116	1213	9.6	
です+けど	61	167	36.5	✗
けど+も	40	150	26.7	✗
で+は	33	217	15.2	
よ+ね	30	147	20.4	
に+は	20	165	12.1	
の+が	19	146	13.0	
て+は	18	168	10.7	

表5 上昇下降調境界音調(L%HL%)における浮き上がり調生起率とアクセント句末2語の関係

句末の2語	FR	L%HL%	%FR	逆接
けれど+も	95	524	18.1	✗
です+が	88	431	20.4	✗
です+けど	59	256	23.0	✗
けど+も	37	206	18.0	✗
です+ね	26	500	5.2	✗
と+か	6	158	3.8	
の+は	6	138	4.3	
に+は	5	64	7.8	
だ+けど	5	30	16.7	✗
です+から	5	63	7.9	

表5に上昇下降調境界音調(L%HL%)について同じ分析を施した結果を示した(この表ではL%HL%の頻度が30に達しないものを除外している)。表4と同一の傾向が生じていることがわかる。

最後に変異の音韻的要因についても簡単に言及しておく。この変異現象については、これまで検討した限り、明瞭な音韻的要因は発見されていない。しかし以下の二つの傾向を指摘することができる。

第一に表6に示すようにアクセント核を含むアクセント句(有核句)とそうでないアクセント句(無核句)を比較すると、無核句において浮き上がり調生起率が若干上昇する傾向が認められる。

第二に表7に示すように上昇調のうち浮き上がり調の句とそうでない句の長さをモーラ数で比較すると、浮き上がり調の生じている句の方が長い傾向が認められる。

表6 アクセント核の有無と浮き上がり調生起率

境界音調	核	FR	H%乃至HL%	%FR
L%H%	有	1016	15205	6.7
	無	523	6886	7.6
L%HL%	有	332	6469	5.1
	無	146	2611	5.6

表7 浮き上がり調とアクセント句の長さ

境界音調	FR 無の発話		FR 有の発話	
	平均	SD	平均	SD
L%H%	6.79	3.84	7.80	3.83
L%HL%	6.49	3.01	7.85	3.48

(単位はすべてモーラ/アクセント句)

以上、アクセント句という単位が重要な役割を果たす研究の例を簡単に紹介した。アクセント句は日本語の韻律研究における最も基本的な単位であり、これをデータ構造中に積極的に採用することによって分析の利便性が高まる。今回再構成したXML文書は、韻律現象に注目した日本語研究のための貴重な言語資源になるだろう。

## 5. おわりに

本稿では日本語話し言葉コーパスの研究用付加情報を統合して表現したXML文書を、特に韻律研究における利便性を高めるためにアクセント句単位を中心に再構成する過程を述べた。再構成したXML文書は近い将来公開する予定である。

## 謝辞:

本研究は文部科学省科学研究費基盤研究(B)「話し言葉コーパスに基づく言語変異現象の定量的分析」(代表者:前川喜久雄, 課題番号16320060)による補助を受けて行った。

## 参考文献:

- [1] 前川喜久雄, 『日本語話し言葉コーパス』の概要, 日本語科学, 15, pp.111-133, 2004.
- [2] 菊池英明他, 「日本語話し言葉コーパスの音声ラベリング」, 音声研究, vol.7, no.3, pp.16-26, 2003.
- [3] 国立国語研究所, 「日本語話し言葉コーパスの構築」, 国立国語研究所報告書124, 2006.
- [4] 川上素, 「文末などの上昇調」, 国語研究, 16, pp.25-46, 1963.