

Internal Use Only

TR-I-0079

連続音声データベースにおける
言語・韻律情報

Integrating linguistic and prosodic information
in a continuous speech database

阿部 匡伸 匂坂 芳典 桑原 尚夫
Masanobu ABE Yoshinori SAGISAKA Hisao KUWABARA

(ATR自動翻訳電話研究所)
ATR Interpreting Telephony Labs.

1989.4

内容梗概

ATRは、各種研究目的に共同利用可能な音声データベースの構築を進めている。これらの音声データベースを様々な観点で有効利用するためには、より多面的かつ詳細な情報を盛り込んでおくことが重要である。本報告では、連続音声データベースに対して付加した形態素属性、係り受け構造等の言語情報、ならびにアクセント句境界、アクセント核位置等の韻律情報について述べる。

1.はじめに

ATRは、各種研究目的に共同利用可能な音声データベースの構築を進めている。これまでに階層的音韻ラベル付けを提案し、ラベル付きの音声データベースを提供してきた^{[1][2]}。また連続音声データベース^{[3][4]}に対しては、新たに基本周波数を追加した^[5]。これらの音声データベースを様々な観点で有効利用するためには、より多面的かつ詳細な情報を盛り込んでおくことが重要である。そこで、我々は連続音声データベース^{[3][4]}(文を読み上げた音声)に対して、言語情報、韻律情報を追加した。本報告では、これらの追加情報について述べる。なお、本データベースを基本周波数データベースと併用する方のために、付録4に基本周波数データベースのデータ形式を掲載した。詳しくは、参考文献[5]参照。

2.言語情報の導入目的

一口に音声の研究と言っても、その研究分野の幅は広い。例えば、音声は単なる音響信号と見なすこともできるし、音声理解の研究対象にもなる。このようなことから、音声データベースはできるだけ広範囲な領域で利用できるように、様々な情報を付加しておくのが望ましい。

音声は人間のコミュニケーション手段であり、意味や文構造との関係があると考えられている。例えば、文構造が基本周波数の変化パターンに影響を及ぼすことはよく知られている^{[6][7]}。従って文や文章の音声データベースにおいては、文構造等の言語情報、及び文音声のアクセント句境界、アクセント核の位置等の韻律情報を導入しておくことが重要であると考えられる。そこで我々は、ATRの連続音声(文を読み上げた音声)データベース^{[3][4]}に新たに言語情報、韻律情報を追加した。

3.付加した言語・韻律情報

文を4つの観点から構成要素に分割し、分割した構成要素に対して属性を与えた。4つの観点とは、音韻、形態素、文節、アクセント句である。以下に各クラスの意味、属性を述べる。

3.1音韻情報

音韻情報を表1に示す。音韻情報は、音声データベースのラベルファイルのうち頻繁に使用すると考えられるものである。詳細な音韻情報が必要な場合には、音声データベースのラベルファイル^[8]を参照する。

3.2形態素情報

形態素情報を表2に示す。また、形態素の品詞属性を表3に、活用型の属性を表4に、活用形の属性を表5に示す。形態素解析は、ATR自動翻訳電話研究所データ処理研究室の方針にそって行った^[9]。品詞の考え方は、付録1を参照のこと。CV音節数とモーラ数との違いは、モーラ数が促音を1つに数え上げることである。

3.3 文節情報と係り受け構造

文節に関する情報には、各文節毎の情報と文節間に決められる情報(係り受け構造)とがある。内容を表6に示す。文節単位への分割は、ATR自動翻訳電話研究所データ処理研究室の方針にそって行った¹⁹⁾。文節の考え方は、付録2を参照のこと。

係り受け構造とは、文節間の関係を木構造で表現する方法である。係り受け構造は人手によって求め、その表記は分離度によって行った。分離度は、隣接する文節間に1つ設定される。その意味は、該当文節が直接係る文節までの文節数である。例を図1に示す。文の係り受け構造(木構造)は、文節の数と文節間の分離度とによって一意に決まる。また文節数と分離度を用いて、存在しえない木構造のチェックも可能である。係り受け構造を分離度によって表記した例を付録3に示す。また、連続音声データベース(503文)のリストを係り受け構造及び分離度付きで別冊付録にまとめた。

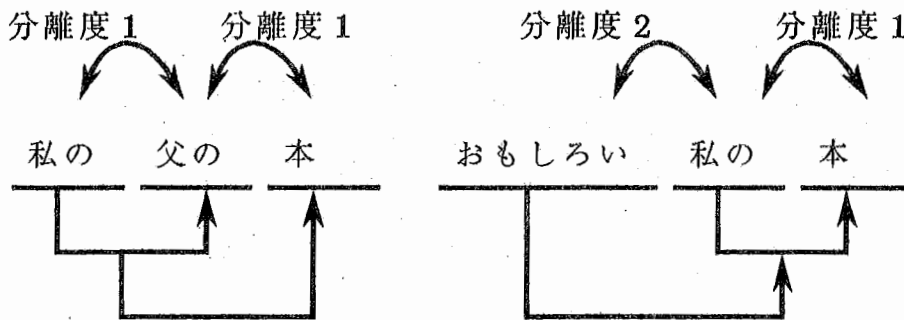


図1 係り受け構造と分離度の関係

3.4 アクセント句情報

アクセント句情報を表7に示す。アクセント句とは、アクセント核を1つだけ含む単位である。アクセント核位置は、聞き取りによって決定した。聞き取りは、言語音声研究の経験を有する1名の女性が行った。形態素、文節、アクセント句で分割した例文を図2に示す。

表1 音韻情報

音韻情報	備考
一文中の音韻数	
音声記号情報1	音声データベースの音声記号層と一致
音声記号情報2	音韻情報のうち、助詞「は」「へ」「を」を「わ」「え」「お」と変換
長母音・融合フラグ	長母音=1 融合=2 その他=0
母音の無声化フラグ	無声化=1 その他=0

表2 形態素情報

形態素情報	備考
一文中の形態素数	
各形態素のリスト	
各形態素のCV音節数	
各形態素のモーラ数	
各形態素の音節数	
各形態素の品詞	種類とコードについては表3参照
各形態素の活用型	種類とコードについては表4参照
各形態素の活用形	種類とコードについては表5参照

表3 品詞

属性	コード
記号	0
形容詞	1
普通名詞	4
サ変名詞	5
代名詞	6
数詞	7
副詞	8
連体詞	9
接続詞	10
感動詞	11
助動詞	12
副助詞	13
接続助詞	14
格助詞	15
終助詞	16
接尾語	17
接頭語	18
補助動詞	19
固有名詞	30
形容名詞	31
本動詞	32
間投詞	33
準体助詞	34
並列助詞	35
係助詞	36

表4 活用型

属性	コード
変則型活用	0
五段活用	1
上一段活用	2
下一段活用	3
サ変活用	4
カ変活用	5
特殊型活用	6

表5 活用形

属性	コード
未然形	0
連用形	1
終止形	2
連体形	3
仮定形	4
命令形	5

表6 文節情報

文節情報	備考
一文中の文節数	
各文節のリスト	
各文節のCV音節数	
各文節のモーラ数	
各文節の音節数	
係り受け構造	

表7 アクセント句情報

アクセント句情報	備考
一文中のアクセント句数	
各アクセント句のリスト	
各アクセント句のCV音節数	
各アクセント句のモーラ数	
各アクセント句の音節数	
アクセント核の位置	アクセント型で表記

	難しい	食事	療法	から	下痢	の	世話	まで	二十	四	時間	介護	の	日々	が	続い	た
	むずかしい	しょくじ	りょうほう	から	げり	の	せわ	まで	にじゅう	よ	じかん	かいご	の	ひび	が	つづい	た
形態素境界																	
文節境界																	
アクセント句境界																	

図2-1 形態素、文節、アクセント句分割の例

	なつかしい	プロペラ	機	で	ふわり	ふわり	と	地球	を	一周	した	ところ	が	すばらし	い
	なつかしい	ぷろぺら	き	で	ふわり	ふわり	と	ちきゅう	を	いっしゅう	した	ところ	が	すばらし	い
形態素境界															
文節境界															
アクセント句境界															

図2-2 形態素、文節、アクセント句分割の例

4. データ形式

連続音声データベースの1ファイル(1文の音声に対応する)に対して、1つの言語・韻律情報ファイルが存在する。言語・韻律情報ファイル名は、音声データベースの命名法に準拠している¹⁰⁾。命名のフォーマットを図3に示す。

話者名: 性別(M=男、F=女)、話者のイニシャル(YI)
タスク: 連続音声データベース、自由発声(_SD_)
文章番号: 文は10のグループに別れており(A~J),
各グループが50文(01~50)からなる。
識別子: 言語・韻律情報データ(INF)

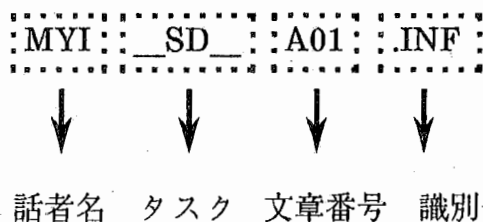


図3. 言語・韻律情報ファイル名のフィールド構成

データは言語・韻律情報ファイル内にASCII形式によって書き込まれている。言語・韻律情報ファイル内は、音韻情報、形態素情報、文節情報、アクセント句情報の4つのクラスに別れている。図4に概要を示す。各クラス間の区切り記号は、"##"である。各クラス内の情報の区切り記号は"#"である。各クラス内の情報の属性は、cr(キャリッジリターン)を区切り記号として書き込まれている。言語・韻律情報ファイルの例を図5に示す。

6. まとめ

連続音声データベースに言語情報を追加した。これによって、音声データベースの利用範囲が広がり、有効利用されることが期待できる。

謝辞 本研究を進める機会を与えてくださった樽松社長に深謝致します。また日頃有益な討論を頂く鹿野室長をはじめとするATR自動翻訳電話研究所研究員の皆様に感謝致します。さらに、形態素解析、文節解析に御協力頂いた、データ処理研究室の森元室長、小倉主任及び篠崎研究技術員に感謝致します。

参考文献

- [1]武田、匂坂、片桐"音声データベース構築のための音韻ラベリング",音講論集,2-5-10(1987.3)
- [2]武田、匂坂、片桐、桑原"音韻ラベルを持つ日本語音声データベースの構築",音声研資,sp87-19(1987)
- [3]磯、渡辺、桑原"音声データベース用文セットの設計",音講論集,2-2-19(1988.3)

- [4]Kuwabara,Takeda,Sagisaka,Morikawa,Watanabe"Construction of a Large-Scale Japanese Speech Database and its Management System",ICASSP'89 (1989.5)
- [5]阿部、桑原"連続音声の基本周波数データベース"、ATRテクニカルレポート、TR-I-0078 (1989.4)
- [6]箱田、佐藤"文音声における音調パタンの諸性質",音声研資,s76-19 (1976.10)
- [7]藤崎、広瀬他"連続音声におけるアクセント成分の実現",音声研資,s84-36 (1984.7)
- [8]武田、匂坂、片桐、阿部、桑原"研究用日本語音声データベース利用解説書"、ATRテクニカルレポート、TR-I-0028 (1988.5)
- [9]水野、篠崎、小倉、吉本"形態素情報利用解説書"、ATRテクニカルレポート、TR-I-0077 (1989.4)

図4. 言語・韻律情報ファイルのデータ形式概要

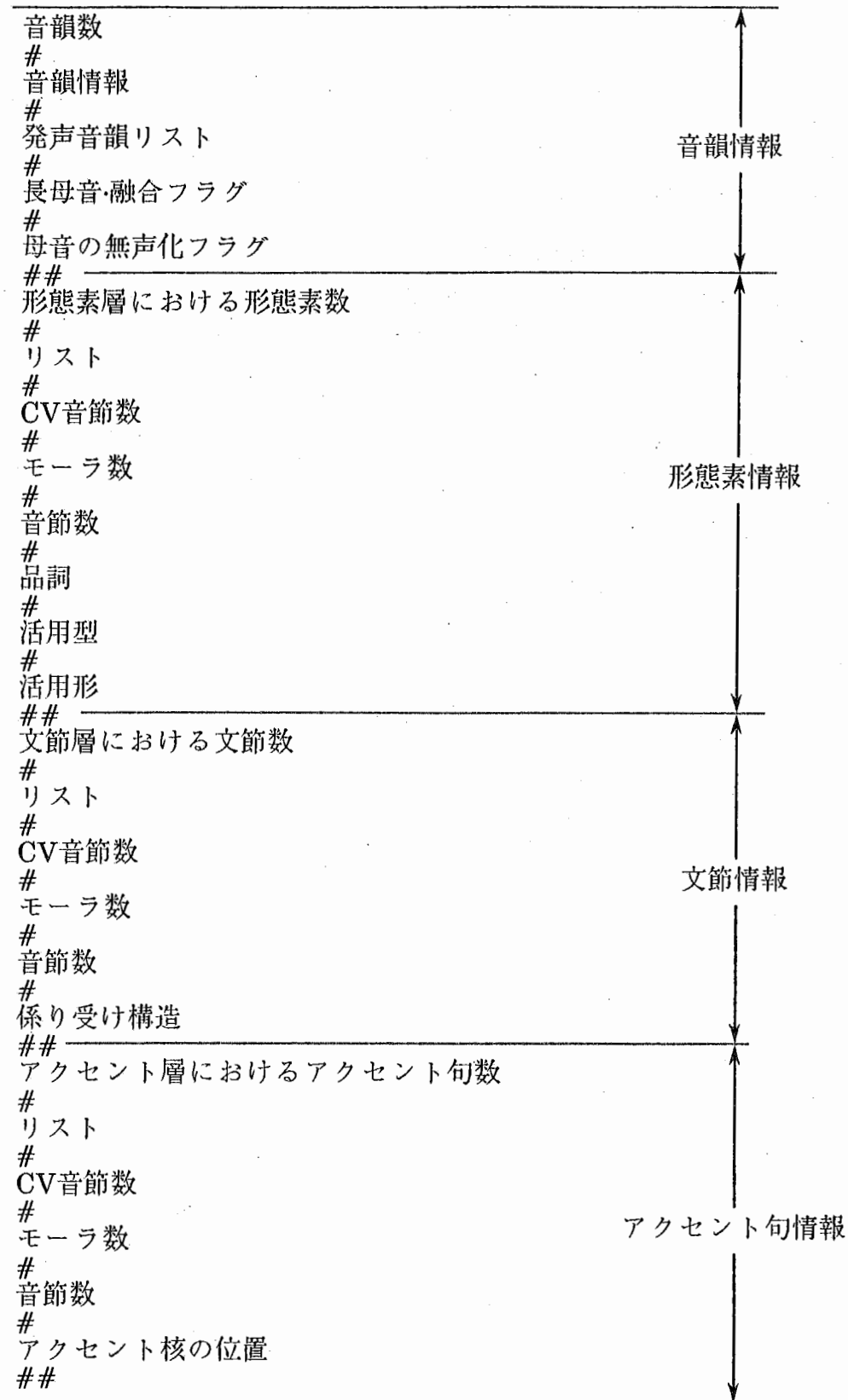


図5. データ形式

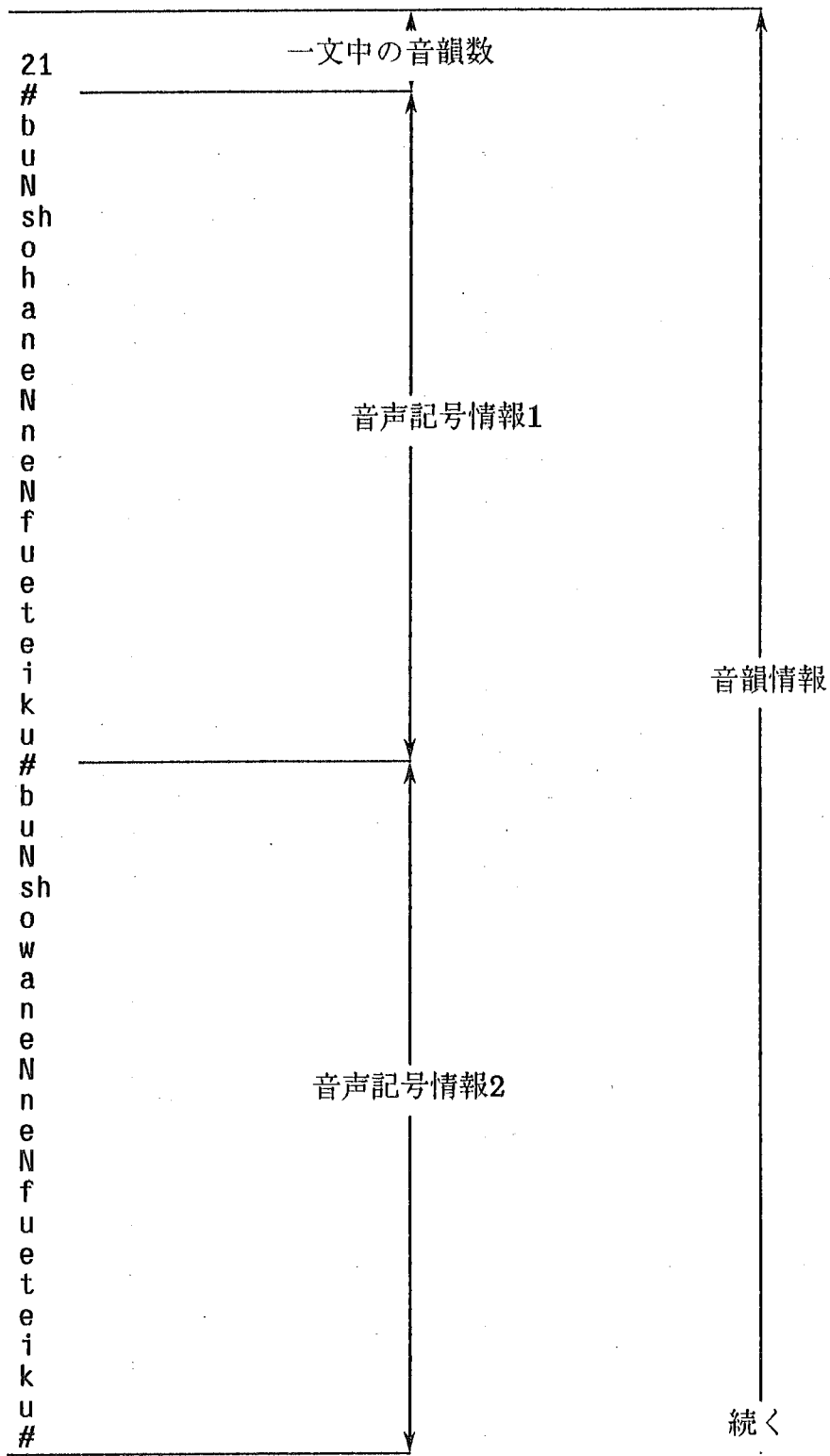


図5. データ形式(続き)

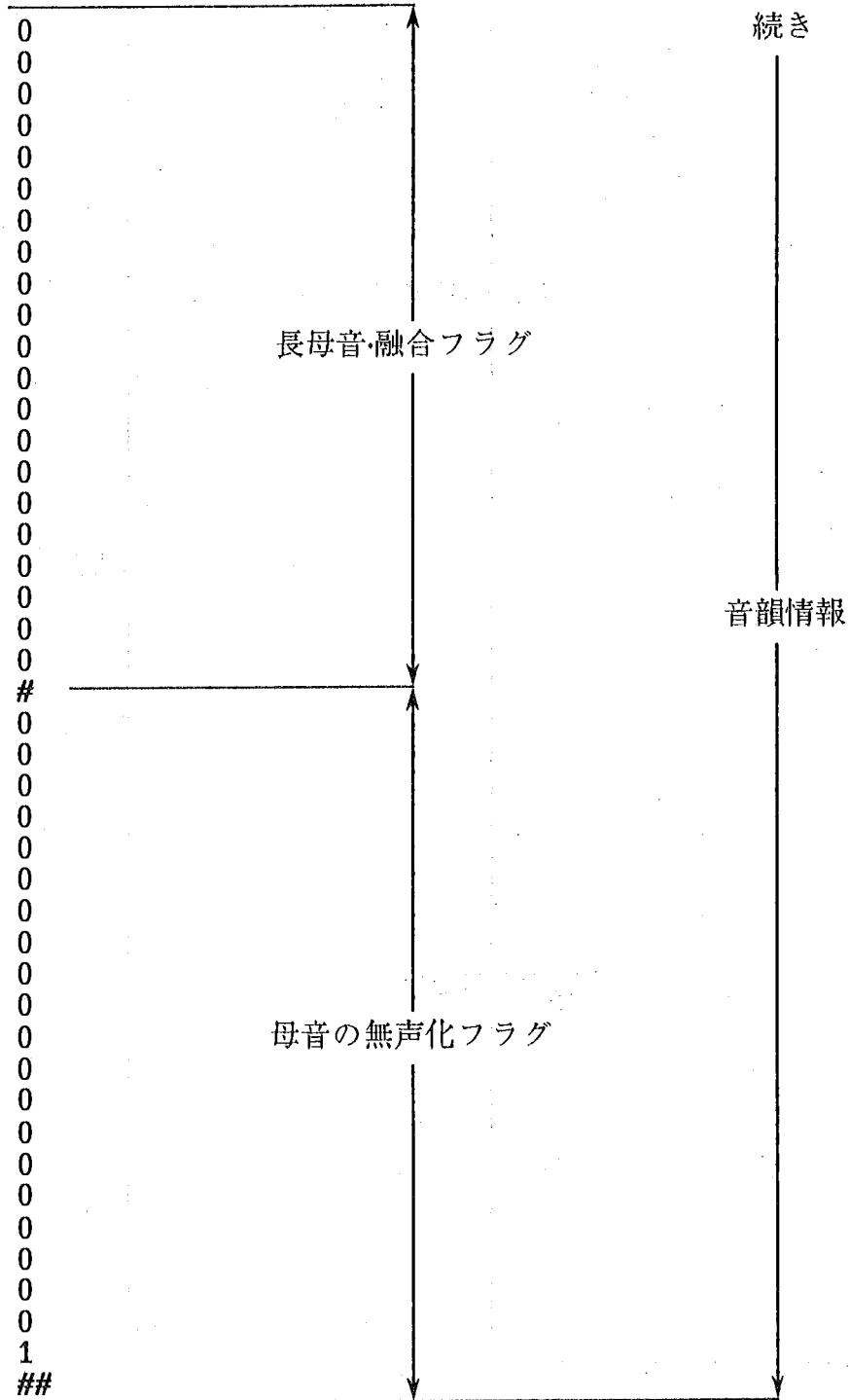


図5. データ形式(続き)

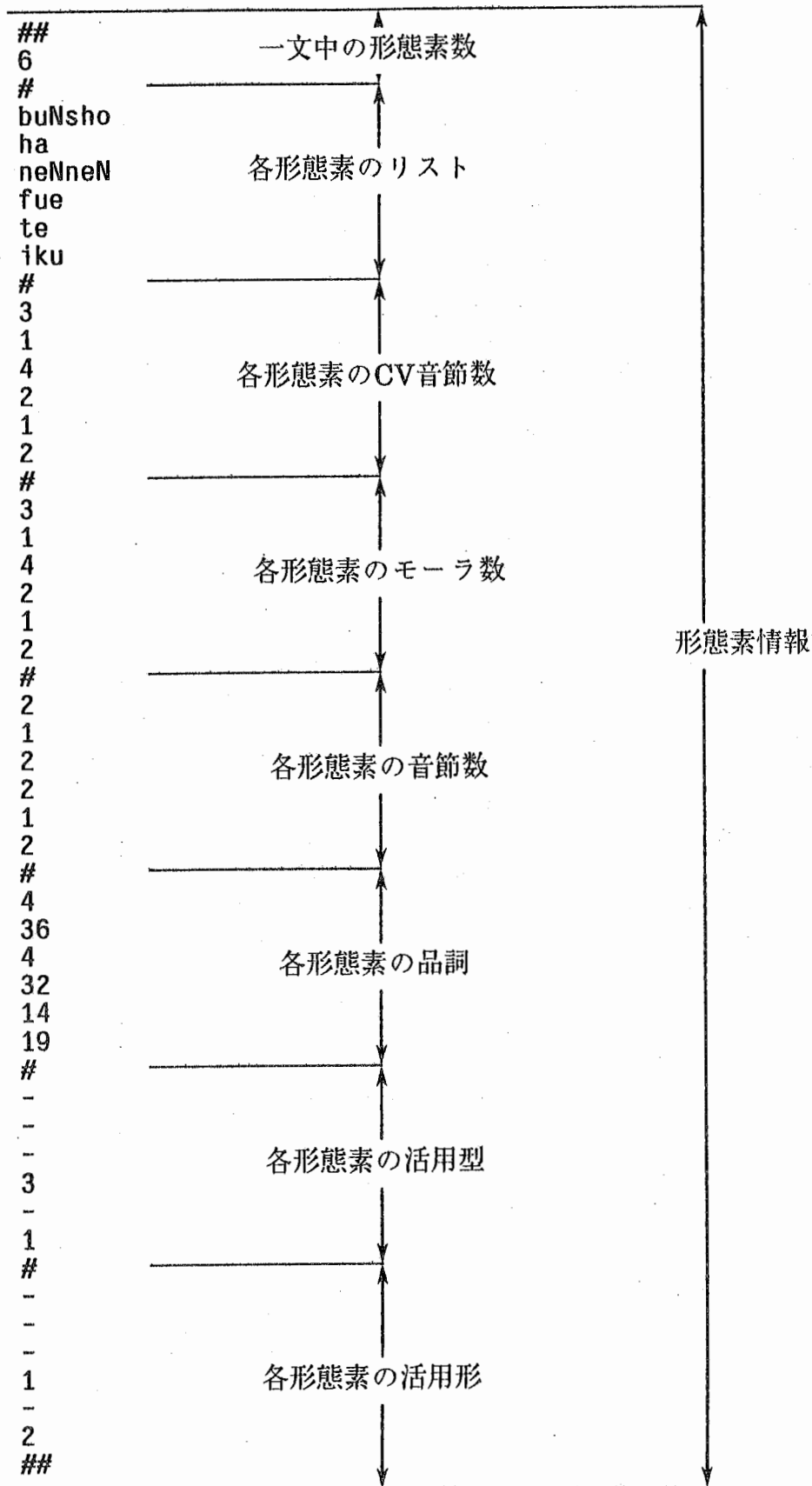
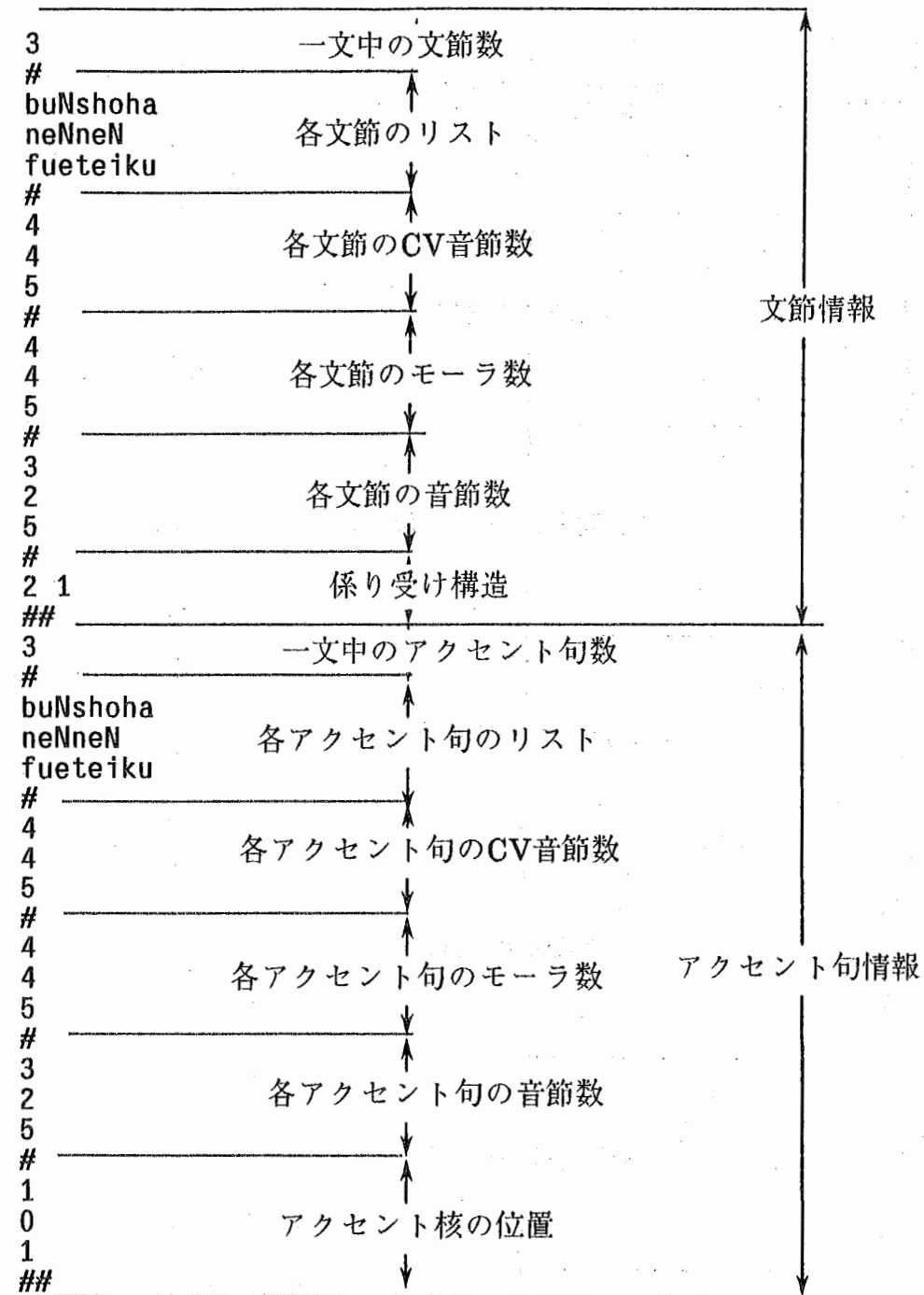


図5. データ形式(続き)



付録1 品詞に対する基本方針⁽⁹⁾

- 1.名詞: 【1.1固有名詞、1.2サ変名詞、1.3形容名詞、1.4普通名詞、1.5数詞、1.6代名詞、1.7名詞その他】

固有名詞、サ変名詞は長単位分割とし、それ以外に関してはできる限り短単位分割とする。最小単位の判断基準としては基本的には新明解を使用しているが、接頭辞・接尾辞を含むものは学研で確認のうえ(更に)分割する。
- 2.動詞: 【2-1.本動詞、2-2.補助動詞、2.3動詞その他】

尊敬表現に関して、独立した判断基準をたてている。動詞の最小単位の判断基準も新明解見出し語としているが、助動詞を含めた形で見出し語となっているものについては辞書記述を採用せず分割する。
- 3.形容詞: 補助形容詞は補助動詞に含めている。辞書内の不統一によるデータのばらつきを抑えるため、一部学研を参考にしてはいるが、基本的には新明解を基準とする。
- 4.副詞: 副詞には名詞との境界が曖昧だという問題があるが、ここではその判断の基準を新明解におくこととしている。ただ新明解には副詞の扱いが不統一な部分があるので、頻出する副詞のパターンをマニュアルで提示し、切り方を統一している。
- 5.連体詞: 「こう、そう、ああ、どう」のつくものは扱いを一本化しているが、それ以外に関しては、判断の基準を新明解におくことにしている。
- 6.接続詞: 不規則な形がよく現れるので、よく出てくるタイプ別に例をのせ、統一的に扱えるよう方針を示している。
- 7.間投詞: (特になし)
- 8.感動詞: 従来の感動詞にとりこぼしや不統一がないかをチェックしている。特に電話会話で種々の感動詞が予想されたため、感動詞の認定基準についてコメントを記している。
- 9.助動詞: 文法的機能語である助動詞に関しては、助詞と同じく、数を限り列挙している。それら以外は原則として助動詞とは認めない。よって、新明解の解釈は採用しない。

10.助詞: 【10-1.格助詞、10-2 準体助詞、10-3 係助詞、10-4 副助詞、
10-5 並立助詞、10-6 接続助詞、10.7 終助詞、10.8.助詞その他】

文法的機能語である助詞に関しては、助動詞と同じく、数を限り
列挙している。それら以外は原則として助詞とは認めない。
よって、新明解の解釈は採用しない。

(助動詞、助詞にとりこぼしが出てきた場合は、学研等を参考にし、「でき
る限り追加を回避する方向で検討する」のを基本方針とする。最終的な決
定はATR担当者がする。)

11.接頭辞: 接尾辞と共に学研を基準とする。これを含む時は、新明解で見出
し語となってもできる限り分割することにする。

12.接尾辞: 接頭辞と共に学研を基準とする。これを含む時は、新明解で見出
し語となってもできる限り分割することにする。

付録2 文節とは何か?⁽⁸⁾

「文節とは何か」について『中学生の国文法の実力』(学友出版)によると以下のように説明されている。

文をくみだしていることばを、意味がわかる範囲内で、できるだけこまかくくぎった一つ一つ。…(途中略)…文を文節に分けるときは、できるだけたくさん「ね」をいれてみるとよい。

しかし、「ね」を挿入できる箇所というのは曖昧であり、判断に個人差が生じる可能性がある。そこで、ここでは形態素解析の結果を利用して、「自立語1個を中心に文節は構成される」と考えることとする。

<基本ルール> 接頭辞*+自立語+接尾辞*+付属語*

ここで、

①「*」は0個以上複数の連続を許すことを表している。

②自立語とは以下の品詞を指す。

名詞(固有名詞,普通名詞,サ変名詞,形容名詞,数詞,代名詞)

本動詞

形容詞

副詞

連体詞

接続詞

感動詞

間投詞

③付属語とは以下の品詞を指す。

補助動詞

助動詞

助詞(格助詞,係助詞,副助詞,並立助詞,接続助詞,終助詞,準体助詞)

<基本ルールに対する例外>

1.自立語が0個という例外

(1)接頭辞+接尾辞で1つの自立語扱いとなるもの。

(例) 幾 日

何 人

2.自立語が2個以上という例外

(2)複数の単語(特に普名詞)の連続が複合名詞として1自立語扱いとなるもの。

(例) 口座 番号

付録 3

係り受け構造及び分離度の例

付録4 基本周波数データフォーマット

基本周波数は、自動抽出後、視察による修正を加えたものである。分析条件を表8に示す。基本周波数データは、基本周期のポイント数で表現されており、c言語の `short integer` 形式で格納されている。修正された基本周期の一部には、修正後の基本周期にマイナス1を掛けてある(どの修正法で行ったかを示すフラグである。詳しくは、参考文献[4]参照)。

基本周波数データベースのファイル名は、音声データベースの命名法に準拠している。命名のフォーマットは、識別子がINFとなること以外は本文図3と同じである。

表8. 分析条件

音声データ	12KHzサンプリング 16bit
分析窓	ブラックマンウインドウ
分析窓長	384ポイント (32msec)
分析シフト	30ポイント (2.5msec)
FFTポイント数	512ポイント
ピッチ探索範囲	男性話者 50~350ポイント (34Hz ~240Hz) 女性話者 24~120ポイント (100Hz ~500Hz)