

TR-IT-0291

## 韻律処理プログラムの改良

佐渡 詩郎 藤澤 謙 Nick Campbell

1999.2.26

### 概要

韻律処理プログラム mor2bou に関して、二重母音処理のバグ修正、母音の無声化処理の C での実装、「中米向け援助」の読みに関するバグ修正、数詞の読み・アクセントに関するバグ修正、ブレークインデックス区切り (BI) の挿入規則の変更を行い、数詞処理に関連して、形態素解析プログラム morph の形態素解析用辞書の修正を行った。これに関連して、gfliter の修正を行った。また、kan2rom の評価を行った。

©ATR Interpreting Telecommunications  
Research Laboratories.

©ATR 音声翻訳通信研究所

# もくじ

1	はじめに	1
2	二重母音処理のバグ修正	3
2.1	二重母音の結合規則	3
2.2	不具合	3
2.3	原因	3
2.3.1	修正ファイル	4
2.4	修正内容	4
2.4.1	新規ファイル	4
2.5	結果	4
2.5.1	実行例	5
3	母音の無声化処理の C での実装	8
3.1	母音の無声化の規則	8
3.1.1	無声子音について	8
3.1.2	母音の無声化の規則	8
3.2	実装	10
3.2.1	入出力	10
3.2.2	新規ファイル・修正ファイル	10
3.2.3	実装方法	10
3.2.4	Perl 版との相違点	11
3.2.5	無声化処理についての指示	11
3.2.6	準無声化処理について	11

3.3	動作試験	11
3.3.1	出力の正当性の確認	12
3.3.2	速度の計測	12
3.3.3	バグとその修正	12
4	「中米向け援助」のアクセントに関するバグ修正	14
4.1	不具合	14
4.2	原因	14
4.3	修正内容	14
4.3.1	修正ファイル	14
4.3.2	ソースプログラムの文字コードの修正	15
4.3.3	バグの修正	15
4.4	動作試験	15
4.5	結果	16
5	数詞の読みに関するバグ修正	18
5.1	助数詞 \$、¥に関するバグ修正	18
5.1.1	不具合	18
5.1.2	原因	18
5.1.3	修正ファイル	18
5.1.4	修正内容	19
5.1.5	動作試験	19
5.1.6	結果	19
5.2	記号「\$」を伴う数詞のアクセントに関するバグ修正	20
5.2.1	不具合	20
5.2.2	原因	20
5.2.3	修正ファイル	20
5.2.4	修正内容	21
5.2.5	結果	21
5.3	「ドル（カタカナ）」を伴う数詞のアクセントの修正	21

5.3.1	不具合	21
5.3.2	原因	22
5.3.3	修正ファイル	23
5.3.4	修正内容	23
5.4	「l\$ です」が「いちです」と読まれるバグの修正	24
5.4.1	不具合	24
5.4.2	原因	24
5.4.3	修正ファイル	24
5.4.4	修正内容	24
5.5	「%」を伴う数詞のアクセントのバグ修正	25
5.5.1	不具合	25
5.5.2	原因	25
5.5.3	修正ファイル	25
5.5.4	修正内容	25
6	形態素解析用辞書の修正	27
6.1	接尾性助数詞「羽」の読みの修正	27
6.1.1	不具合	27
6.1.2	原因	27
6.1.3	修正ファイル	28
6.1.4	修正内容	28
6.1.5	結果	28
6.2	「万一」の読みの修正	28
6.2.1	不具合	28
6.2.2	原因	29
6.2.3	修正ファイル	29
6.2.4	修正内容	29
6.2.5	結果	29
6.3	その他	29

6.3.1	修正ファイル	29
6.3.2	修正内容	29
<b>7</b>	<b>BIの変更</b>	<b>30</b>
7.1	問題点	30
7.2	修正ファイル	30
7.3	修正内容	30
<b>8</b>	<b>C-STAR II 対応版のバグ修正</b>	<b>31</b>
8.1	概要	31
8.2	修正ファイル	31
8.3	修正内容	31
<b>9</b>	<b>gfilter の修正</b>	<b>32</b>
9.1	gfilter 出力での BI の変更	32
9.1.1	概要	32
9.1.2	修正ファイル	32
9.1.3	修正内容	32
9.2	gfilter の無声化処理の C 版への切り替え	33
9.2.1	概要	33
9.2.2	修正ファイル	33
9.2.3	修正内容	33
<b>10</b>	<b>kan2rom の評価</b>	<b>34</b>
10.1	評価対象	34
10.2	評価方法	34
10.3	評価結果	35
<b>11</b>	<b>まとめ</b>	<b>36</b>
<b>12</b>	<b>謝辞</b>	<b>37</b>

<b>13 付録</b>	<b>39</b>
13.1 評価対象文と kan2rom の出力 . . . . .	40
13.2 good.txt . . . . .	56
13.3 CVS 履歴 . . . . .	59
13.4 数詞処理 . . . . .	67
13.4.1 流れ . . . . .	67
13.4.2 数詞 . . . . .	68
13.4.3 数字と単位数字のアクセント結合 . . . . .	70
13.4.4 助数詞処理 . . . . .	71
13.4.5 数字と特殊助数詞のアクセント結合 . . . . .	76
13.4.6 音便処理 . . . . .	76

# 第 1 章

## はじめに

第 2 研究室が開発した多言語音声合成システム CHATR を TTS(Text To Speech; テキスト読み上げ) で用いる場合、kan2rom と組み合わせて使用する。kan2rom は、漢字仮名混じり文をブレイクインデックス (以下 BI) 区切りやアクセントマークを伴ったローマ字に変換するプログラムであり、形態素解析プログラム morph と、韻律処理プログラム mor2bou の二つから成り立っている (図 1.1)。

mor2bou に関しては、<http://www.itl.atr.co.jp/~fujisawa/morph/bugs/aopen.html> でバグ報告がされているが、このうち、以下の部分を改良した。

- 二重母音の結合処理 (/ei/→/ee/, /ou/→/oo/) が行われていないバグを修正した
- 母音の無声化処理は Perl で別モジュールとして実装され、フィルタプログラムとして用いられているが、これを C で実装し mor2bou に組み込んだ
- 「中米向け援助」のアクセント位置がおかしいバグを修正した
- 数詞処理に関連したバグを修正した
- BI の挿入規則を変更した
- MATRIX で用いる場合の数詞処理のバグを修正した

無声化処理と BI の挿入規則変更に関連して、gfliter を修正した。また、数詞処理に関連して形態素解析プログラム morph の形態素解析用辞書を修正した。kan2rom の性能評価を目的として、修正後の kan2rom による評価を行った。以下、第 2 章では二重母音の結合処理のバグ修正について報告し、第 3 章では母音の無声化処理の C での実装について報告し、第 4 章では「中米向け援助」のアクセント位置がおかしいバグの修正について報告し、第 5 章で数詞処理に関連したバグの修正について報告し、第 6 章では形態素解析用辞書の修正について報告し、第 7 章では BI の挿入規則の変更について報告し、第 8 章では C-STAR II 対応版のバグ修正について報告し、第 9 章では gfilter の修正について報告し、第 10 章では kan2rom の評価について報告する。

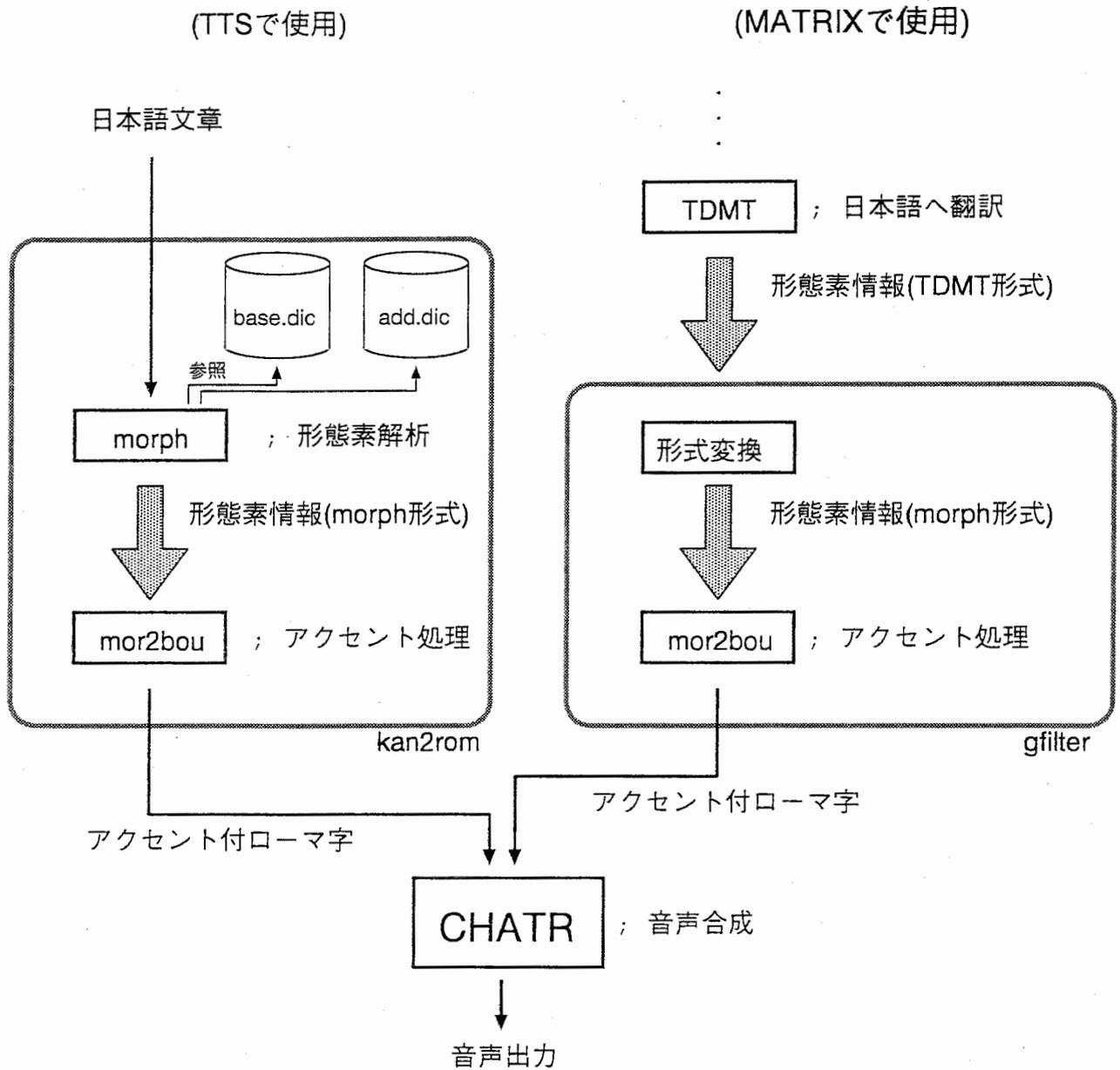


図 1.1: システム構成の概略

## 第 2 章

### 二重母音処理のバグ修正

#### 2.1 二重母音の結合規則

二重母音の発音に次の規則（文献[1]）がある（表 2.1）。

- エ行に続くイは /e/ と発音される。（ /ei/ → /ee/ ）
- オ行に続くウは /o/ と発音される。（ /ou/ → /oo/ ）

ただし、単語が結合されて上記の母音の組み合わせになった場合、この規則は適用されない（表 2.2）。

#### 2.2 不具合

2.1で述べた二重母音の結合規則が実行されず、次のようになる。

- イが常に /i/ と発音される（ /ei/ → /ei/ ）
- ウが常に /u/ と発音される（ /ou/ → /ou/ ）

#### 2.3 原因

韻律処理プログラム mor2bou のモジュールの一つである、PROSODY/cal\_accbou/calaccent.c のソースコード中に、韻律処理構造体配列の連結を行う関数 irt\_cat() が記述されている。この関数では、単語結合時に発生する二重母音 /ei/、/ou/ について、処理 /ei/ → /ee/ および /ou/ → /oo/ が行われることを抑止する目印として、分離記号「”」（ダブルクォーテーション）を挿入している。韻律処理出力を確認したところ、出力中に分離記号「”」が正常に付加されていることが確認された。

calaccent.c による韻律処理出力は、PROSODY/cal\_accbou/conv.c により、アクセント記号「'」（シングルクォーテーション）および分離記号「”」等でタグ付けされた、JIS X 201 カナ（半角カナ）による出力形式に変換される。PROSODY/cal\_accbou/hfr.c により、この形式は、JIS X 201 カナからローマ字綴りによる ASCII 列に変換され、mor2bou の最終的な出力となる。

分離記号「”」は、conv.c に記述されている関数 str\_chk() で除去されている。しかし、母音の書き換え規則 /ei/→/ee/ および /ou/→/oo/ に関しては、どの部分でも実装されていない。したがって、最終的な出力として、/ei/→/ee/ および /ou/→/oo/ の処理が行われておらず、また、irt\_cat() での処理も、反映されない。

このことから、母音の書き換え規則を実装すれば良いとわかった。

### 2.3.1 修正ファイル

修正したファイルは以下の通りである。

PROSODY/cal\_accbou/conv.c[1.5]

## 2.4 修正内容

conv.c に記述されている関数 str\_chk() では、

- 入力列中に含まれる「"」、「^」（ハット）を取り除く

という処理が行われている。しかし、この処理以前に、二重母音処理を行うコードが存在しないため、不具合となっている。

そこで、str\_chk() を変更し、二重母音処理を行うコードを追加した。このコードは、JIS X 201 カナを対象に処理を行うように記述した。

### 2.4.1 新規ファイル

以下の新規ファイルを作成した。

PROSODY/cal\_accbou/kana\_def.h[1.1] (JIS X 201 カナのコードを定義)

## 2.5 結果

正しく形態素解析が行われたものについて、2.1 で述べた規則と合致する結果が得られた。

### 2.5.1 実行例

入力文の例(1)、および、それに対する韻律処理出力の中間出力（デバッグ時のみ）、最終出力を示す(表 2.3)。入力文の下線部 5 箇所が、2.1で示した規則により、書き換えの対象となる部分である。/ei/ または /ou/ が、それぞれ /ee/ または /oo/ に書き換えられていることが確認できる。

次に、入力文の例(2)、および、それに対する韻律処理出力の中間出力（デバッグ時のみ出力可能）、最終出力を示す(表 2.4)。入力文の下線部 5 箇所は、2.1で示した規則により、書き換えの対象から除外される部分である。入力文に /ei/ および /ou/ が含まれるが、どちらも書き換えられていないことが確認できる。

表 2.1: 母音の発音が変化する例 (太字が対象部分)

語句	かな表記	発音
発酵	はっこう	ハッコー
羽毛布団	うもうぶとん	ウモーブトン
強烈	きょうれつ	キョーレツ
王様	おうさま	オーサマ
どうしよう	どうしよう	ドーシヨウ
清潔	せいけつ	セーケツ
成功	せいこう	セーコー
低下	ていか	テーカ
抵抗	ていこう	テーコー
泥炭	でいたん	データン
米国	べいこく	ベーコク
晴天	せいてん	セーテン

表 2.2: 母音の発音が変化しない例 (太字が対象部分)

語句	かな表記	発音
小ウサギ	こうさぎ	コウサギ
舞の海	まいのうみ	マイノウミ
歩いていた	あるいていた	アルイテイタ
している	している	シテイル

表 2.3: プログラム修正後の実行結果例 (1)

入力文	こうして、水泳の名人がようやく到着した。
中間出力	0 コウシテ            こうして nmd1 0 スイエイノ        水泳の 0 メイジ'ンガ'      名人が nmd2 0 ヨウヤク           ようやく nmd1 0 トウチャク        到着した - シタ decl
最終出力	kooshite 1 suieeno meeji'Nga 2 yooyaku 1 toochaku shita 5

表 2.4: プログラム修正後の実行結果例 (2)

入力文	彼は子ウサギの世話で手一杯だ。
中間出力	0 カ'レワ 彼は nmd1 0 コ'ウ'サキ'ノ 子ウサギの 0 セワ'テ' 世話で 0 テ'イ'ッパイタ' 手一杯だ decl
最終出力	ka'rewa 1 kou'sagino sewa'de tei'ppaida 5

## 第 3 章

### 母音の無声化処理の C での実装

mor2bou では、母音の無声化規則 (文献 [2]) に基づいた無声化フィルタが Perl で実装されているが、処理に時間がかかるとされているので、同様の処理を C で実装し、mor2bou に組み込んだ。

#### 3.1 母音の無声化の規則

##### 3.1.1 無声子音について

以下で定義する母音の無声化の規則で、無声子音の定義を用いるが、この定義は、Perl 版無声化フィルタの実装で用いられている定義と同一とする (表 3.1)。

表 3.1: 無声子音とその分類

分類	子音
破裂音	/k/ /t/ /p/
摩擦音	/sh/ /s/ /f/ /h/
破擦音	/ts/ /ch/

##### 3.1.2 母音の無声化の規則

母音 /i/、/u/ に対する無声母音をそれぞれ /i/、/u/ とする (ただし、実装上は、それぞれ /I/、/U/ とした)。実装した母音の無声化の規則は次の通りであり、文献 [2] に基づいている。

無声子音に挟まれた母音 /i/、/u/ は無声化する。

(ex. /ashika/ → /ashika/, /eNpitsu/ → /eNpitsu/ )

ただし、以下の条件が満たされなければならない (図 3.1)。

- アクセント核がある場合、無声化しない（準無声化）  
(ex. /chi'shiki/ → /chi'shiki/)
- 直前の母音が無声化している場合、無声化しない（準無声化）  
(ex. /shikifuku/ → /shikifuku/)
- 同種の無声摩擦音に挟まれた場合、無声化しない（準無声化）  
(ex. /susumu/ → /susumu/)

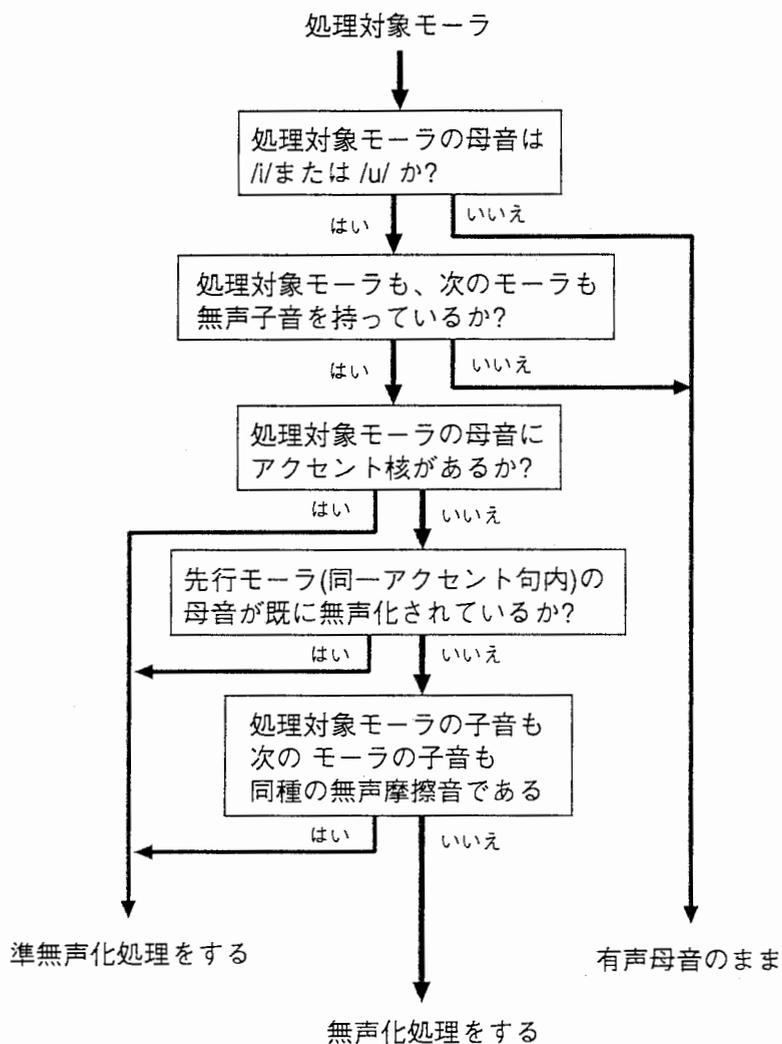


図 3.1: 母音の無声化処理の流れ図

また、助動詞「ます (/masu/)」「です (/desu/)」の /u/ は、必ず無声化する。

## 3.2 実装

### 3.2.1 入出力

入力は無声化処理のされていないローマ字綴りによる文字列、出力は無声化処理のされたローマ字綴りによる文字列とした。

Perl 版にならい、無声子音 /i/, /u/ の出力形式は、それぞれ I, U とした (表 3.2)。

表 3.2: 母音と無声母音の表記

母音	入出力形式	無声母音	出力形式
/a/	a	—	—
/i/	i	/i/	I
/u/	u	/u/	U
/e/	e	—	—
/o/	o	—	—

### 3.2.2 新規ファイル・修正ファイル

作成した新規ファイルは以下の通りである。

PROSODY/cal\_accbou/unvoice.c[1.1] (母音の無声化処理を実装)

修正したファイルは以下の通りである。

PROSODY/cal\_accbou/hfr.c[1.2] (unvoice() の呼び出し)

PROSODY/cal\_accbou/mor2bou.c[1.5] (無声化スイッチの追加)

PROSODY/cal\_accbou/hfr.c[1.3] (無声化スイッチの追加)

PROSODY/cal\_accbou/unvoice.c[1.2] (バグ修正)

### 3.2.3 実装方法

3.1.2に示した無声化の規則に基づいて実装を行った。

unvoice.c に関数名 unvoice() として無声化の処理を実装した。変更ファイルとして、JIS X 201 カナからローマ字綴りによる文字列への変換を行うモジュール hfr.c を変更し、関数 hfr() から unvoice() を呼び出すようにした。

また、kan2rom のスイッチとして '-u' (無声化処理を行う) があったが、これをやめ、'-U' (無声化処理を行しない) を設けた。

### 3.2.4 Perl版との相違点

Perl版との主な相違点は次の通りである。

- Perl版では入力を一旦モーラ単位に切り分けていたが、C版ではこの切り分け処理を行っていない
- Perl版は mor2bou とは独立したプログラムであり、標準入出力を介して mor2bou とのデータ授受を行っていたが、C版は mor2bou に組み込んだ

### 3.2.5 無声化処理についての指示

kan2rom は、デフォルトでは無声化処理をせず、コマンド行からのオプション `-u` が存在する場合、無声化処理を行うよう指示するようになっていた。これを変更し、デフォルトで無声化処理を行い、mor2bou へのコマンド行からのオプション `-U` が存在する場合、無声化処理を行わないよう指示するようにした。

### 3.2.6 準無声化処理について

ソースコード `unvoice.c` で定義されているコンパイル時のマクロ定数 `UNVOICE_SEMI` を書き換えることにより、準無声化処理に関する指示を行うことができるよう実装した(表 3.3)。ただし、

表 3.3: 準無声化処理に関する指示

UNVOICE_SEMI	準無声化処理に関する指示
0	無声化しない
1	無声化処理と同様の処理をせよ
2	準無声化として処理せよ

現在、CHATR は準無声化処理に対応していないため、出力形式としても、準無声化処理が用意されていない。このため、マクロ変数 `UNVOICE_SEMI` は、準無声化処理を無声化処理と同様に扱うように設定した。

## 3.3 動作試験

プログラムの動作の確認を行った。入力の文章には、音素的にバランスのとれた 503 文章 (以下、ATR503 文章) を元に kan2rom で作成したローマ字のファイル (以下 503.romaji) を用いた。実行ファイルは、`unvoice.c[1.1]` を元に `gcc unvoice.c -DDEBUG.UNVOICE` として作成した `a.out` を C 版無声化フィルタ、`unvoice/unvoice.pl` を Perl 版無声化フィルタとした。

### 3.3.1 出力の正当性の確認

プログラムの正当性についての確認内容は次の通りである。

- 準無声化フラグを無声化に指定して 503.romaji に無声化処理を行い、その結果を Perl 版の出力と照合した。
- 準無声化フラグを準無声化に指定して、503.romaji を対象に無声化処理を行い、その結果を人手による作業でルールと比較した。

いずれの動作試験でも、プログラムが正常に動作することが確認された。

### 3.3.2 速度の計測

as52.itl.atr.co.jp 上で速度の計測を行った。一旦、C での無声化処理を呼び出す C 版の無声化フィルタを作成し、ATR503 文章を KDD kan2rom で BI を伴うローマ字に変換したファイルを対象に、Perl 版無声化フィルタ、C 版無声化フィルタの両方で処理を行い、これにかかる時間を time コマンドを用いて計測した。503.romaji を単位として、それを 1~3 回繰り返したものを入力とし、/dev/null に対して出力を行った(表 3.4)。Perl 版では単位文あたりに約 27 秒、オーバヘッ

表 3.4: 実行時間の計測結果

プログラム	503.romaji	×2	×3
Perl 版	28.5~29(秒)	56	83
C 版	0.14~0.15(秒)	0.24	0.34

ドに 1.5~2.0 秒、C 版では単位文あたりに約 0.1 秒、オーバヘッドに 0.04~0.05 秒を所要していることがわかる。

### 3.3.3 バグとその修正

文末の「です/ます」が無声化されない、というバグが発見された。

無声化処理を、最初にフィルタプログラムとして作成したが、この時は、関数 unvoice() の引数として渡される、処理対象の文字列が文全体を含むことを仮定していた。

しかし、mor2bou への組み込みを行った場合、PROSODY/cal\_accbou/hfr.c からの呼び出しでは、無声化処理がアクセント句ごとに行われる。

呼び出しを、文単位に行うよう、呼び出し場所を変更することを検討したが、アクセント句を結合して文単位で一つにまとめる処理は、最終的な出力の書き出しによって行われているだけであった。

そこで、`unvoice.c`を変更し、文末の「です/ます」の処理について、BIを参照せずに、アクセント句ごとに行うようにした [1.2]。

## 第 4 章

### 「中米向け援助」のアクセントに関するバグ修正

#### 4.1 不具合

「中米向け援助 (/chuubeemukēe'Njo/)」に、誤ったアクセントづけ /chuubeemuke'eNjo/ がされる。

#### 4.2 原因

PROSODY/cal\_accbou/calaccent.c [1.6] で、複合語の韻律処理が行われている (関数 acc\_mark() 中 第 1423~1427 行)。calaccent.c では、複合語の韻律処理の結果で長音の最後にアクセントが来た場合 (「アー」)、アクセントを 1 モーラ前に移動する (「ア'ー」) ようになっている。

「中米向け」と「援助」が複合された「ちゅうべいむけえんじょ」が、この処理により長音と判定され、アクセントが移動してしまう。

#### 4.3 修正内容

##### 4.3.1 修正ファイル

修正したファイルは以下の通りである。

PROSODY/cal\_accbou/calaccent.c [1.7] (不具合の修正 および 文字コードの修正)

PROSODY/cal\_accbou/calaccent.c [1.7] (文字コードの修正)

### 4.3.2 ソースプログラムの文字コードの修正

ソースプログラム `calaccent.c` [1.6] は MS 漢字コード (シフト JIS 漢字コード) を用いて記述されていたが、次の点

- コメント中に JIS X 201 カナ (半角カナ) が使用されている。
- MS 漢字コード (2 バイト) の直前に必ず `0x92` が挿入されている。
- JIS X 201 カナ (1 バイト) の直前に必ず `0x89` が挿入されている。

が、作業上不便であると思われた。そこで、JIS X 201 カナ (1 バイト) はコメント内でしか使われていないため、これを漢字コード (2 バイト) を用いた同一のカタカナに置き換えることにより、ソースプログラム中で JIS X 201 カナが使われないようにした。また、MS 漢字コードの直前に挿入されている `0x92` を取り除いた [1.7]。

### 4.3.3 バグの修正

`calaccent.c` で、形態素の複合処理を行う `irt_cat()` から呼び出されている関数 `kana_cat()` がある。`kana_cat()` では、二重母音の読みに関する情報として、形態素の区切りに「"」 (ダブルクォーテーション) が挿入されている。[1.6] では、二つの形態素の複合について、後方の形態素の第 1 モーラの母音が /i/ または /u/ の場合にのみ「"」が挿入されるようになっていたが、これを変更して、第 1 モーラの母音の種類に関わらず「"」が挿入されるようにした [1.7]。

## 4.4 動作試験

デバッグ用に、同一の入力に対する修正前のプログラムの出力と修正後のプログラムの出力の差分を出力する `csh` 用のスクリプトファイルを `~xssado/for_debug/debug_mor2bou` として作成した。これを用いて次の動作試験を行った。

1. 「中米向け援助 (/chuubeemukē'Njo/)」が、正しく読まれることを確認した。
2. ATR503 文章を対象に、修正前と修正後の出力を比較した。  
(対象ファイル `~xssado/textdata/atr503.txt`)
3. `www.asahi.com` からダウンロードした「本日の朝刊 (1 月 26 日付)」を対象に、修正前と修正後の出力を比較した。  
(対象ファイル `~xssado/textdata/asahi.com.99-01-26.txt`)

1. は、修正されたことが確認できた。2. は、修正前と修正後で違いはみられなかった。3. は、修正前に誤って出力されて、修正後に正しく出力されたものが 5 件確認された (表 4.1)。

## 4.5 結果

同一の入力に対し、間違いから正解への変化が確認され、正解から間違いへの変化は確認されなかった。このことから、「中米向け援助」に関するバグは修正されたと考えてよいと思われる。

表 4.1: 「本日の朝刊」で確認された修正結果

元の文	修正前 修正後
巨人の渡辺恒雄オーナーは	kyojiNno watanabe tsuneo'oonawa kyojiNno watanabe tsunee'onaawa
横綱曙について	yokozuna'akebononi tsuite yokozunaa'kebononi tsuite †
時津風理事長（元大関豊山）は	togitsUfuuriji'choo 3 moto'oozeki toyoyama 3 wa togitsUfuuriji'choo 3 motoo'ozeki toyoyama 3 wa ‡
今年の箱根駅伝で	kotoshino hakone'ekideNde kotoshino hakonee'kideNde
箱根駅伝2区で	hakone'ekideN 1 ni'kude hakonee'kideN 1 ni'kude

†……完全には正しくない。/yokozuna 1 akebono/ と出力されるのが良い。

‡……/mo'to o'ozeki 1 toyoyama/ と出力されるのが良い。

## 第 5 章

### 数詞の読みに関するバグ修正

#### 5.1 助数詞 \$、¥に関するバグ修正

##### 5.1.1 不具合

「\$1 です」が「いちどるいちです」と読まれてしまう。

##### 5.1.2 原因

PROSODY/cal\_accbou/num\_kana.c (CVS revision 1.5) では、数詞および助数詞について、表記から読みへの変換を行っている。KDD morph による形態素解析では、「\$」と「1 です」は複数の形態素に分割される。分割された個々の形態素は、mor2bou に入力された際に、韻律処理構造体配列の要素に変換される。その後、韻律処理構造体配列の連続した要素にまたがる数詞は、PROSODY/cal\_accbou/num\_kana.c に記述されている関数 num\_yomi() で、再び、韻律処理構造体配列の一つの要素にまとめられる。その際に、韻律処理構造体配列の接続処理を行うが、このための指示がおかしかった。そのため、韻律処理構造体配列の走査で、「\$1」の「1」の部分が2重に走査され、「\$1 です」が「いちどるいちです」と読まれていた。

##### 5.1.3 修正ファイル

修正したファイルは次の通りである。

PROSODY/cal\_accbou/num\_kana.c[1.6] (修正)

PROSODY/cal\_accbou/num\_kana.c[1.7] (emacs 用タグにミスがあったので修正)

PROSODY/cal\_accbou/num\_kana.c[1.8] (修正方法を変更して修正)

#### 5.1.4 修正内容

num\_kana.c を変更し、数詞が韻律処理構造体配列の連続した要素にまたがる場合の、終点の指示を変更した。また、ソースコードのコーディング法に関して、保守性の面で問題が感じられたので、関数 num\_yomi() に関して、書法の面での書き換えを行った。

なお、一旦修正を行ったソースコードを登録したが、後日、修正方法の不具合に気づき、再度の修正を行った。

#### 5.1.5 動作試験

以下のテキストを用いて動作試験を行った。

1. 問題となる数詞及び助数詞による表現を人手により作成したもの  
~xssado/for\_debug/1doru/suushi.txt
2. www.nikkei.co.jp からダウンロードした新聞記事  
~xssado/for\_debug/1doru/nikkei.txt
3. www.asahi.com からダウンロードした新聞記事  
~xssado/textdata/asahi.com.99-01-26.txt  
~xssado/textdata/asahi.com.99-01-27.txt  
~xssado/textdata/asahi.com.99-01-28.txt  
~xssado/textdata/asahi.com.iraq.txt  
~xssado/textdata/asahi.com.suma.txt  
~xssado/textdata/asahi.com.wakayama.txt  
~xssado/textdata/asahi.com.takuhai.txt

1. は、バグ症状が確実に出るものばかりで作ったものである。2. は、数詞および助数詞を使った表現が多く見られるテキストとして用いた。3. は、修正による悪影響が出ていないことを確認するために用いた。

動作試験は、1. については、修正後の mor2bou の出力を検証した。また、1.~3. について、修正前と修正後の2種類のバイナリによって吐き出された、同一の入力に対する出力の差分を検証した。

また、~xssado/for\_debug/1doru/difference.txt として、テキストファイルにこの差分の検証結果をまとめた。これを印刷したものを、巻末に付録として添付する。

#### 5.1.6 結果

「\$1 ドル」が「いちどるいち」と読まれる不具合は修正されていることが確認された。

また、「3:25 です」が /sa'Nji ni'juu go'fuN ni'juu go 3 de'su/ と読まれる不具合についても、修正されていることが確認できた。

## 5.2 記号「\$」を伴う数詞のアクセントに関するバグ修正

### 5.2.1 不具合

記号「\$」を伴う数詞に付与されるアクセント記号の位置が不適切であり、「\$1(/ichi'doru/)」が /ichido'ru/ と読まれる。

### 5.2.2 原因

助数詞を伴う数字列のアクセント記号を付与した読みを求める関数 `num_kana()` が、`num_kana.c` に記述されている。これを呼び出す形で、助数詞 \$ を伴う数詞に読み、および、アクセント記号を付与する関数が、`doru_yomi()` であり、同ファイル中に記述されている。

`doru_yomi()` が `num_kana()` を呼び出す際に、`num_kana()` の第4引数として、数詞が助数詞を伴う際のアクセントの移動型を指定している (文献 [1], func-64)。このアクセントの移動型は、

1. 助数詞が漢字一文字、もしくは、カタカナ／ひらがな 2 モーラ (引数: 100)
2. 助数詞のアクセント型が平板型／頭高型 (引数: 101)
3. 助数詞のアクセント型が中高型 (引数: 100 + モーラ数)

と分類されている (括弧内はアクセントの移動型として渡される引数の値)。また、1. には、「カタカナ／ひらがな 2 モーラの場合は、長音、撥音となる場合、および、辞書に記された特定単語」という制限がある。この「辞書」について調べたが、文献 [1] では、一見すると「辞書」が何を指すのか不明である (実際に何を指すかについては 5.3.2 参照)。

このことから、おそらく、`doru_yomi()` のコーディングの際に、「辞書」が何を指すのか不明のまま、「ドル」のアクセントがアクセントの移動型として 2. の 101 を指定したと思われる。しかし、2. に該当する場合に付与されるアクセント位置は不適切なものであり、適切なアクセントの付与を行うためには、アクセントの移動型として 1. を指定しなければならない。

### 5.2.3 修正ファイル

修正したファイルは次の通りである。

PROSODY/cal\_accbou/num\_kana.c[1.9]

## 5.2.4 修正内容

num\_kana.c に記述されている doru\_yomi() を変更した。num\_kana() の呼び出しの際に第4引数として指定されている助数詞のアクセントの移動型を、「平板型／頭高型」(101) から「漢字一文字、もしくは、カタカナ／ひらがな2モーラ以下」(100) に変更した。

## 5.2.5 結果

修正の結果、ドルの読みとして期待されるものが得られるようになった(表5.1)。

表 5.1: 助数詞ドルへのアクセント付与の修正結果

入力	出力 (修正前)	出力 (修正後)
1 ドル	/ichido'ru/	/ichi'doru/
2 ドル	/nido'ru/	/ni'doru/
3 ドル	/saNdo'ru/	/sa'Ndoru/
4 ドル	/yoNdo'ru/	/yo'Ndoru/
5 ドル	/godo'ru/	/go'doru/
6 ドル	/rokudo'ru/	/roku'doru/
7 ドル	/nanado'ru/	/nana'doru/
8 ドル	/hachido'ru/	/hachi'doru/
9 ドル	/kyuudo'ru/	/kyu'udoru/
10 ドル	/juudo'ru/	/ju'udoru/
11 ドル	/juuichido'ru/	/juuichi'doru/
100 ドル	/hyakudo'ru/	/hyaku'doru/
1000 ドル	/seNdo'ru/	/se'Ndoru/

## 5.3 「ドル (カタカナ)」 を伴う数詞のアクセントの修正

### 5.3.1 不具合

助数詞「ドル (カタカナ)」を伴う数詞に付与されるアクセント記号の位置が不適切であり、「1ドル /ichi'doru/」が /ichido'ru/ と読まれる。

### 5.3.2 原因

mor2bou では、記号「\$」と助数詞「ドル」に対して、異なる処理が与えられている。「\$」の場合、num\_kana.c 中で、数字処理を行う num\_yomi() から doru\_yomi() が呼び出され、ここで、記号「\$」を伴う数詞に読みが与えられている。「ドル」が数詞に後続した場合、KDD\_morph による形態素の解析ととして、助数詞として出力される。mor2bou は、数詞とそれに後続する助数詞に対して、cal\_accent.c 中に記述されている acc\_unit() で、助数詞のアクセント接続型を判定し、判定されたアクセント型を引数の一つとして num\_kana() を呼び出している。

数詞に助数詞が後続した場合のアクセント処理は文献 [1] で記述されているが、一般助数詞は、一旦

1. 漢字1字、もしくは、カタカナ・ひらがな2モーラ以下の場合  
ただし、カタカナ・ひらがな2モーラ以下の場合、次の条件

- 長音・撥音により2モーラとなっている
- 辞書に記された特定単語

のいずれかを満たさなければならない

2. 条件1. に該当せず、助数詞のアクセント型が、平板型、もしくは、頭高型
3. 条件1. に該当せず、助数詞のアクセント型が、中高型

と分類される。

ここで、「辞書に記された特定単語」であるかどうかの判定は、ソースコード cal\_accent.c[1.7] を参照した結果、韻律処理構造体配列中、助数詞のアクセント接続情報上位1バイトが、2である、という条件を満たす場合、「辞書に記された特定単語」としている。ただし、アクセント接続情報上位1バイトは、形態素解析処理によって付与された情報であり、形態素解析用辞書に記述されているものと同一のものである。

しかし、形態素解析用辞書 base.dic[1.9] に記述されている助数詞「ドル」の項目で、

- アクセント接続情報 上位1バイト …… 0
- アクセント接続情報 下位1バイト …… 0
- アクセント複合情報 …… 2

となっている。つまり、形態素解析用辞書に記述されている助数詞「ドル」の、アクセント接続情報上位1バイトは2になっておらず、2になっているのはアクセント複合情報である。アクセント接続情報上位1バイトに関して、文献 [1] は、

- 0 無処理
- 1 「の」の特殊処理 (尾高型に低くつく処理)
- 2       "       (尾高型にする処理)

と記述するのみであるが(p.14)、アクセント複合情報については、

- 0 なし
- 1 サ変動詞の境界を1つ前に移動
- 2 助数詞に低くつく2モーラカタカナ

と記述している。また、「ミリ」等の類似した性質を持つ単語でも、同様に、アクセント複合情報が2になっていることから、「辞書に記された特定単語」であるかどうかの判定として、アクセント接続情報上位1バイトではなく、アクセント複合情報を用いるのが適切であると考えられる。

また、同じく `acc_unit()` 中の、「カタカナ・ひらがな2モーラ以下」を判定する部分で、助数詞がカタカナまたはひらがなであることの判定に `strncmp` を用いている。これは、`strncmp` の比較が `unsigned char` で行われることを期待しており、処理系が `char` で比較している場合は誤った判定を行う。

```

/* カタカナ、ひらがな2モーラ以下 */
if( (strncmp(kanji,"あ",2)>=0 && strncmp(kanji,"ん",2)<=0)
    || (strncmp(kanji,"ア",2)>=0 && strncmp(kanji,"ヶ",2)<=0) ){

```

図 5.1: `strncmp` を用いたひらがな・カタカナの判定 (`calaccent.c`[1.7] より)

### 5.3.3 修正ファイル

修正したファイルは次の通りである。

PROSODY/cal\_accbou/cal\_accent.c[1.8]

PROSODY/cal\_accbou/inritu.h[1.4]

### 5.3.4 修正内容

`inritu.h` で、カタカナ・ひらがなである条件の判定用マクロ `CHK_KANA` および `CHI_HIRA` が記述されている。これも `strncmp` を用いているので、用いないように書換えた。また、そのままでは、

条件が「一」（棒引き）をカタカナ・ひらがなとしていないので、これを修正し、「一」はカタカナ、および、ひらがなであるとした(図 5.2)。

cal\_accent.c 中に記述されている関数 acc\_unit() を変更した。「カタカナ・ひらがな 2 モーラ以下」を判定する部分で、助数詞がカタカナまたはひらがなであることを判定するために、strcmp を用いて範囲を判定していたが、CHK\_KANA および CHI\_HIRA を用いるように変更した。また、「辞書に記された特定単語」であるかどうかの判定に、アクセント接続情報上位 1 バイトが用いられていたが、アクセント複合情報を用いるようにした。

## 5.4 「1\$ です」が「いちです」と読まれるバグの修正

### 5.4.1 不具合

「1\$ です」が「いちです」と読まれてしまう。

### 5.4.2 原因

通常の表記である「\$1」と異なり、「1\$」とは普通書かない。このためか、入力が「1\$」であった場合に、処理が不適切に行われる。

### 5.4.3 修正ファイル

PROSODY/cal\_accbou/num\_kana.c[1.10]

### 5.4.4 修正内容

num\_kana() に記述されている、数字列の処理部分である、に「\$1」の形式で記述された、数詞と助数詞の組み合わせを追加した。

## 5.5 「%」を伴う数詞のアクセントのバグ修正

### 5.5.1 不具合

記号「%」を伴う数詞（「1% (/ippaase'ntto/)」など）に、間違ったアクセント位置 /ippaa'sentto/ が付与される。

### 5.5.2 原因

ルール「漢字1字、もしくは、カタカナ・ひらがな2モーラ以下の場合」（5.3.2参照）に、実装上、「%」が該当してしまう。

### 5.5.3 修正ファイル

PROSODY/cal\_accbou/calaccent.c[1.9]

PROSODY/cal\_accbou/inritu.h[1.5]

### 5.5.4 修正内容

「漢字1字」というルールの実装部分 (acc\_unit(), cal\_accent.c) を修正した。2バイト文字1字であれば「漢字1字」と見なしていたが、表5.2を含む記号 (JIS X 208 非漢字) は漢字と見なさないようにした。ただし、角度を表す「°」は、「度」と同様と見なし、「漢字1字」とするようにした。

表 5.2: 漢字1字と見なさない2バイト文字

%	パーセント
\$	ドル
Å	オングストローム
‰	パーミル
¢	セント
£	ポンド
'	インチ
"	フィート
℃	摂氏度

```

/* ひらがな (シフト JIS の文字コードで比較) */
#define      CHK_HIRA(str)                                \
/* - */          ( ( ((unsigned char *)str)[0] == 0x81    \
                    && ((unsigned char *)str)[1] == 0x5b)  \
/* あ - ん */   || ( ((unsigned char *)str)[0] == 0x82    \
                    && ((unsigned char *)str)[1] >= 0x9f  \
                    && ((unsigned char *)str)[1] <= 0xf1  ) )

/* カタカナ (シフト JIS の文字コードで比較/半角も) */
#define      CHK_KANA(str) /* ア <= x <= ケ or 半角'ラ'<=x<=半角'ン' */ \
/* - */          ( ( ((unsigned char *)str)[0] == 0x81    \
                    && ((unsigned char *)str)[1] == 0x5b)  \
/* 第1byte */   || ( ((unsigned char *)str)[0] == 0x83    \
/* 及、第2byteの */ && ((unsigned char *)str)[1] >= 0x40  \
/* 範囲を比較 */ && ((unsigned char *)str)[1] <= 0x96  ) \
/* x>= 半角'ラ' */ || ( ((unsigned char *)str)[0]>=0xa6    \
/* x<= 半角'ン' */ && ((unsigned char *)str)[0]<=0xdd ) )

```

図 5.2: strncmp を用いないひらがな・カタカナの判定 (inritsu.h[1.5] より)

## 第 6 章

### 形態素解析用辞書の修正

#### 6.1 接尾性助数詞「羽」の読みの修正

##### 6.1.1 不具合

接尾性助数詞「羽」は、1羽、3羽など特別な場合に「イッパ」「サンバ」と読まれることがある場合を除くと、「は」と表記され、「ワ」と読まれる。

しかし、mor2bou では、正しくない読みが付加される場合がある (表 6.1)。付加される読みは、

表 6.1: 「羽」の読みの不具合

入力	付加される読み	正しい読み
1羽	イッパ	イチワ, イッパ
2羽	ニハ	ニワ
3羽	サンバ	サンバ
10羽	ジュッパ	ジュウワ, ジュッパ
100羽	ヒヤッパ	ヒヤクワ, ヒヤッパ
1000羽	センハ	センワ, センバ

「ワ」とするべきところを「ハ」としている以外はうまく読まれているように見える。

##### 6.1.2 原因

mor2bou は、文献 [1] で記述されたルールに基づくが、文献 [1] では助数詞「羽」が元々の読みとして「ワ」を持つものと設定している。しかし、辞書に助数詞「羽」が、「ワ」および「ハ」の2種類の読みで設定されていたため、助数詞「羽」の読みとして「ハ」が選択され、「羽」の読みに関するルールがうまく働かなかったと考えられる。

なお、ここで用いられているのは「ハ」で始まる助数詞（「班」など）のためのルールであり、「3羽」で「羽」が「バ」と読まれたのは、文献[1]で助数詞が「羽」の場合に設定されている、特別な場合のためのルールに、実装上、該当してしまっただからだと思われる。

### 6.1.3 修正ファイル

修正したファイルは以下の通りである。

KANJI/base.dic[1.8]（形態素解析用辞書）

### 6.1.4 修正内容

このことから、形態素解析用辞書 base.dic から助数詞「羽（は）」を取り除いた。

### 6.1.5 結果

この結果、助数詞「羽」の読みとして、文献[1]の仕様に基づくものが付加されるようになった（表 6.2）。

表 6.2: 「羽」の読みの不具合の修正結果

入力	付加される読み（変更後）
1羽	イチワ
2羽	ニワ
3羽	サンバ
10羽	ジュウワ
13羽	ジュウサンバ
100羽	ヒャクワ
1000羽	センワ

## 6.2 「万ー」の読みの修正

### 6.2.1 不具合

「万ー（まんいち）」は「マンイチ」と読まれるが、フレーズ「万ーのことを考えると」で、「万ー（まんいち）」に「イチ」という読みが付加される。

## 6.2.2 原因

形態素解析時に「万一」が数詞とされていた。このため、mor2bou「万一」が「0万 + 1」と解釈され、この結果得られた「1」に対する読み「イチ」が、「万一」に読みとして付加される。

## 6.2.3 修正ファイル

修正したファイルは次の通りである。

KANJI/base.dic[1.9] (形態素解析用辞書)

KANJI/add.dic[1.13] (形態素解析用辞書)

## 6.2.4 修正内容

base.dicに、「万一」が属性510(副詞「に」)として登録されていたので、これを属性500(副詞一般)に変更した。

add.dicに、属性100(名詞一般)として「万一」を登録した。

## 6.2.5 結果

「万一のことを考えると」という場合にも、「万一」が「イチ」ではなく「マンイチ」と読まれるようになった。

## 6.3 その他

### 6.3.1 修正ファイル

修正したファイルは次の通りである。

KANJI/add.dic[1.13] (形態素解析用辞書)

### 6.3.2 修正内容

人名「昭宣 (あきのり)」を登録した。

## 第 7 章

### BI の変更

#### 7.1 問題点

CHATR のバグリストに、「スペース区切り (BI 無し) のローマ字列を合成すると正しく target F0 を予測しない」がある (<http://www.itl.atr.co.jp/~fujisawa/CHATR/bugs/BI/index.html>)。この問題への対応方法の一つとして、「kan2rom を必ず BI 付きで出力するように修正する」ことが挙げられている。

#### 7.2 修正ファイル

修正したファイルは次の通りである。

PROSODY/cal\_accbou/mor2bou.c[1.8]

#### 7.3 修正内容

mor2bou.c に記述されている mor2bou\_text\_out() を変更し、BI を伴わない空白区切りが存在した場合、強制的に BI(1) を挿入するようにした [1.7]。

一度エンバグしてしまった (BI が多く入りすぎる場合がある) ので、それを修正した [1.8]。

## 第 8 章

### C-STAR II 対応版のバグ修正

#### 8.1 概要

kan2rom が C-STAR II で用いられる際に、「(数字) です」の形式での入力に対し、次の仕様が与えられている (表 8.1)。

- 入力の数字は漢数字のみである
- 桁読みの必要がある場合、十、百、千等が入力として含まれる

#### 8.2 修正ファイル

修正したファイルは次の通りである。

```
PROSODY/cal_accbou/num_kana.c[1.11]
```

#### 8.3 修正内容

マクロ定数 MATRIX を定義してコンパイル (gcc なら -DMATRIX) した場合のソースコードを変更した。韻律処理構造体に含まれる数詞の漢字表記に「何」「数」「幾」「余」「千」「百」「十」のいずれかが含まれる場合、数詞を「分離読み」ではなく桁読みとするようにした。

表 8.1: C-STAR II での数字読み上げ

入力	読み上げ形式
二三四	ニイサンヨン
二百三十四	ニヒャクサンジュウヨン

## 第 9 章

### gfilter の修正

mor2bou の修正に関連して、gfilter を修正した。<sup>1</sup>

#### 9.1 gfilter 出力での BI の変更

##### 9.1.1 概要

CHATR は、スペース区切り (BI 無し) のローマ字列を合成すると正しく target F0 を予測しない。mor2bou を修正し、この問題に対処したが (第 7 章)、出力修正箇所が mor2bou.c であるため、gfilter が出力する BI を変更するためには、gfilter のソースコードを変更する必要がある。

##### 9.1.2 修正ファイル

修正したファイルは次の通りである。

```
gfilter/src/tdmt_mor2bou.c[1.3]
```

##### 9.1.3 修正内容

tdmt\_mor2bou.c に記述されている mor2bou\_text\_out() を変更し、BI を伴わない空白区切りが存在した場合、強制的に BI(1) を挿入するようにした。ただし、gfilter は誤って BI を多く入れる場合があるようなので、BI が余分に出力される場合は、最も深い統語境界を持つ BI を 1 つだけ出力するようにした。

---

<sup>1</sup>なお、gfilter は酒井 慎介がバグ修正を担当しており、この修正のみ自分が行った。これ以外の gfilter のバグ修正については、文献 [3] で報告されている。

## 9.2 gfilter の無声化処理の C 版への切り替え

### 9.2.1 概要

gfilter からは Perl 版の無声化フィルタが呼び出されていたが、C 版で実装した無声化フィルタを呼び出すようにした。

### 9.2.2 修正ファイル

修正したファイルは次の通りである。

```
gfilter/src/makefile[1.4]
gfilter/src/gfilter.c[1.5]
morph/PROSODY/cal_accbou/unvoice.c[1.3]
```

### 9.2.3 修正内容

gfilter.c 中に記述されている、Perl 版無声化フィルタを呼び出す関数 unvoice() を削除した。

プロトタイプ宣言

```
static void unvoice( char * )
```

を削除し、新たにプロトタイプ宣言

```
extern void unvoice( char * )
```

を記述した。これは unvoice.c に記述されている unvoice() を呼び出すためのものである。

unvoice.c に記述されている unvoice() を修正した。mor2bou はアクセント句ごとに unvoice() を呼び出すが、gfilter は文ごとに unvoice() を呼び出す。このため、文ごと、アクセント句ごとのどちらで呼び出されても、文末の /desu/、/masu/ の無声化を行うようにした。

## 第 10 章

### kan2rom の評価

#### 10.1 評価対象

朝日新聞、日経新聞から抽出された 100 文 (samples.txt)。1994 年 1 月、6 月、11 月の日経新聞から 90 文章、および、最近の www.asahi.com から集めた新聞記事から 10 文章。

#### 10.2 評価方法

これを kan2rom で CHATR の入力形式に変換したもの (アクセント句数 1424) を、人手により標準語アクセントを基準として正解・不正解を判定した。この判定結果を元に、次式で示すアクセント正解率、よみ正解率を求めた。

$$\text{アクセント正解率} = \frac{\text{アクセント句数} - \text{間違ったアクセントが付与された句数}}{\text{アクセント句数}}$$

$$\text{よみ正解率} = \frac{\text{アクセント句数} - \text{間違った読みが付与された句数}}{\text{アクセント句数}}$$

評価結果を 10.3 に記述しする。評価対象文、kan2rom による出力、および、誤りの内容を巻末に付録として添付する。

### 10.3 評価結果

評価の結果は次の通りである。

- 100 文中、アクセントも読みも正しく付与できたもの …… 68 文
- 100 文中、読みは正しく付与されたもの …… 87 文
- アクセント誤り …… 25 文中、30 箇所 (34 句)
- よみ誤り …… 12 文中、13 箇所 (18 句)
- アクセント正解率 …… 97.60%  $\left( = \frac{1424 - 34}{1424} \right)$
- よみ正解率 …… 98.73%  $\left( = \frac{1424 - 18}{1424} \right)$

## 第 11 章

### まとめ

mor2bou に関して、

- 二重母音の結合処理が行われていないバグの修正
- 母音の無声化処理を C で実装し、組み込み
- 「中米向け援助」のアクセント位置がおかしいバグの修正
- 数詞処理に関連したバグの修正
- BI の挿入規則の変更

を行い、morph の形態素解析用辞書を修正した。また、これに関連して gfilter の修正を行った。まだ解決されていないバグとして、主なものに

- 「元」、「前」、「米」などの接頭辞がアクセント複合してしまう
- 接尾辞「型」がつく単語が、アクセント複合しない
- 名詞が続くと 1 語としてしまう

が挙げられる。また、数詞処理についても若干のバグが残っている。

(<http://www.itl.atr.co.jp/~fujisawa/morph/bugs/aopen.html> 参照)

kan2rom の評価の結果は良好であり、kan2rom は、テキスト読み上げについて、実用的な性能を発揮できると思われる。しかし、不具合もあることから、今後も若干の修正作業が必要になるだろうと思われる。

## 第 12 章

### 謝辞

本作業をすすめるにあたり、あたたかく見守ってくださった Nick Campbell 室長、ならびにあたたかい御指導と適切な助言を頂きました藤澤 謙研究員に心より感謝します。また、さまざまな有益な助言を頂いた ATR 音声翻訳研究所第二研究室の皆様に深く感謝します。

最後になりましたが、貴重な実習の機会を与えてくださった指導教官の増山 繁教授に心より感謝します。

1998 年 2 月  
佐渡 詩郎

## 参考文献

- [1] 上妻 真史, “韻律処理マニュアル”, システム科学研究所, (1994)
- [2] 佐藤 大和, “規則による音声合成の研究”, 北海道大学 博士論文, (1987)
- [3] 酒井 慎介, “言語フィルターの改善”, ATR 音声翻訳通信研究所, TR-IT-0290, (1999)

## 第 13 章

### 付録

次頁より、

- 10章で述べた kan2rom の評価における、評価対象の文章と kan2rom による出力、および、その誤り部分
- 本報告書で報告した作業に関連したファイルの CVS 履歴
- 5.1.5節で述べた good.txt; 数詞に関するバグ修正の検証結果
- mor2bou における数詞処理についての記述

を添付する。

### 13.1 評価対象文と kan2rom の出力

細川首相は九四年の年頭に当たって記者会見し、中長期的な政策指針「二十一世紀ビジョン（細川私案）」を発表した。

hosokawa 1 shUshoowa 2 kyu'ujuu 1 yone'Nno 1 neNtooni 1 atatte 1 kIshaka'ikeN 1 shi 1 chuuchookItekina 1 seesakUshI'shiN 3 ni'juu 1 isse'ekibijoN 3 hosokawashi'aN 3 o 1 happyoo 1 shIta 5

軽自動車メーカー最大手のスズキと第二位のダイハツ工業は軽自動車の部品を共通化することで基本合意した。

keejidooshame'ekaa 1 saio'oteno 1 su'zukIto 1 daini'ino 1 daihatsUko'ogyoowa 1 keejido'oshano 1 buhiNo 1 kyootsuuka 1 suru 1 kotode 2 kIhoN 1 gooi 1 shIta 5

su'zukIto 1        アクセント誤り (suzukIto 1)  
daini'ino 1        アクセント誤り (da'i 1 ni'ino 1)  
kIhoN 1 gooi 1    アクセント誤り (kIhongo'oi 1)

麻布向け以外の不良債権も含めた全体の償却額は千四百億円前後にのぼる見込みで、信託銀行としては過去最大の不良債権処理になる。

azabumuke 1 i'gaino 1 furyoosa'ikeNmo 1 fUku'meta 1 zeNtaino 1 shookyaku'gakuwa 2 se'N 1 yoNhyakuo'kueN 1 ze'Ngoni 1 noboru 1 mikomide 2 shiNtakugi'NkootoshItewa 2 ka'ko 1 saidaino 1 furyoosaikeNsho'rini 1 naru 5

政府の税制調査会（首相の諮問機関、加藤寛会長）は年金制度改革に合わせて年金税制のあり方を見直し、課税強化を目指す方針だ。

se'efuno 1 zeeseechoosa'kai 3 shUshoono 1 shimoNkI'kaN 3 ka'too 1 hi'roshi 1 kaichoo 3 wa 1 neNkiNseedoka'ikakuni 1 awa'sete 1 neNkiNze'eseeno 1 arika'tao 1 minaoshi 3 kazee 1 kyo'okao 1 meza'su 1 hooshiNda 5

鮮明な動画像も含めた大量の情報を高速で処理するので、通信だけでなく放送用の回線としても使える。

seNmeena 1 dooga'zoomo 1 fUku'meta 1 tairyoono 1 joohooo 1 koosokude 1 sho'ri 1 surunode 2 tsuushiN 1 dakede 1 naku 1 hoosooyoono 1 kaiseNtoshIte'mo 2 tsUkaeru 5

surunode 2        アクセント誤り (suru'node 2)

全国の自治体で、建設・土木会社への職員の天下りに一定の制限を加えたり、組織的なあっせんを廃止する動きが広がってきた。



mata 1 shiNyooki'Nkoya 1 shiNyooku'miaino 1 te'Nponi 1 kaka'ru 1 koteeshIsa'Nzeeno 1  
hIkazeeso'chimo 1 haishi 1 shItai 1 kaNga'ede 1 reNritsu 1 yo'tootono 1 chooseo 1 iso'gu  
5

系列ディーラーが主体となり、第一弾として最大手の東京日産自動車販売が二月中に専門会社を  
設立し、四月にも販売を開始する。

keeretsudi'iraaga 1 shUtaito 1 nari 1 daiichi'daNtoshIte 1 saio'oteno 1 tookyoo 1 nissaNji-  
dooshaha'Nbaiga 1 nigatsUchuuni 1 seNmoNga'ishao 1 setsuritsu 1 shi 1 shigatsu'nimo 2 haN-  
baio 1 kaishi 1 suru 5

daiichi'daNtoshIte 1 アクセント誤り (da'i 1 ichi'dantoIshite 1)

製紙最大手の日本製紙は十四日、販売代理店（一次卸）の数を四月以降、約三分の二まで段階的  
に削減する方針を決めた。

seeshi 1 saio'oteno 1 niho'N 1 seeshiwa 2 ju'uyokka 3 haNbaidairi'teN 3 ichijio'roshi 3 no 1  
ka'zuo 1 shigatsu 1 i'koo 3 ya'ku 1 saNbuNnoni'made 1 daNkaitekini 1 sakugeN 1 suru 1  
hooshiNo 1 kimeta 5

niho'N 1 seeshiwa 2 アクセント誤り (nihoNse'eshiwa 2)

両社は共同で食品を開発、味の素が生産して、ダイエー系列の小売店、外食店に供給する。

ryo'oshawa 2 kyoodoode 2 shokUhiNo 1 kaihatu 3 ajino'motoga 1 seesaN 1 shIte 1  
daieeke'eretsuno 1 kouri'teN 3 gaishokU'teNni 1 kyookyuu 1 suru 5

ただ、適用条件が厳しいため、同制度の認定企業はまだ十五件にとどまっており、通産省は条件  
を緩和することにした。

ta'da 1 tekiyoojo'okeNga 1 kibishi'i 1 tame 3 do'oseedono 1 niNteeki'gyoowa 2 ma'da 1 ju-  
ugo'keNni 1 todoma'tte 1 ori 1 tsuusa'Nshoowa 1 jooke'No 1 kaNwa 1 suru 1 koton 1 shIta 5

現在、二万二千人弱の従業員を大規模な希望退職の募集や関係会社への出向・転籍などで九七年  
度末までに一万七千人まで絞り込む。

ge'Nzai 3 nima'N 1 niseNniNjakuno 1 juugyo'oiNo 1 daiki'bona 1 kiboota'ishokuno 1 boshu-  
uya 1 kaNkeega'ishaeno 1 shUkkoo 3 teNsekina'dode 2 kyu'ujuu 1 nananeNdo'matsumadeni 1  
ichima'N 1 nanaseNniNmada 1 shiborikom 5

niseNniNjakuno 1 アクセント誤り (niseNniNja'kuno 1)

三十代半ばのAさんは数年間の日本での生活を振り返り、「結局、自分は会社のマスコットだっ  
たんだ」という苦い思いにかられる。



kokkaiwa 2 ni'juuyokka 1 go'go 3 saNiN 1 hoNka'igide 2 hIketsu 1 saretu 1 seejikaikakUkaN-  
reNho'oaNno 1 toriatsUkai'o 1 megutte 1 yoya'tooga 1 hoNkakUtekina 1 sesshooni 1 ha'itta 5

住宅対策では住宅金融公庫の融資戸数を六万戸程度上積みする方向で大蔵、建設両省が最終調整  
に入った。

juutakUta'isakudewa 1 juutakUkiNyuuko'okono 1 yuushIkosu'uo 1 rokumaNkoteedouwazumi  
1 suru 1 hookoode 2 ookura 3 keNsetsuryo'oshooga 1 saishuuchu'oseeni 1 ha'itta 5

米国に比べ遅れていた日本のマルチメディア市場の環境整備が大きく進むものとみられる。

beekokuni 1 kurabeoku'rete 1 ita 1 nihoNno 1 maruchimediashi'jooono 1 kaNkyoo 1 se'ebiga 1  
o'okIku 1 sUsumu 1 mono'to 1 mirare'ru 5

kurabeoku'rete 1                      アクセント誤り (kurabe 1 okurete'ita 1)

同時に今国会成立に向け、「両院協議会の状況を踏まえ、自民党の河野総裁にトップ会談を呼び  
掛けたい」との考えを明らかにした。

doojini 1 ko'Nkokkai 1 seeritsuni 1 muke 3 ryooiNkyoogi'kaino 1 jookyooo 1 fuma'e 3 jimiNtoono  
1 koonoso'osaini 1 toppUka'idaNo 1 yobikaketa'i 3 to'no 1 kaNga'eo 1 aki'rakani 1 shIta 5

koonoso'osaini 1                      アクセント誤り (ko'ono 1 soosaini 1)

細川首相は二十八日、連立政権の最大課題であった政治改革実現にメドをつけたことで、今後の  
政権運営の基盤を固めることになろう。

hoso'kawa 1 shUshoowa 2 ni'juuhachinichi 3 reNritsUse'ekeNno 1 saidaika'daide 1 atta 1 see-  
jika'ikaku 1 jitsugeNni 1 me'doo 1 tsUke'ta 1 kotode 1 koNgonono 1 seekeNu'Neeno 1 kibaNo 1  
katameru 1 kotoni 1 naro'o 5

細川連立政権が命運をかけた政治改革関連四法案は二十九日午後、衆参両院本会議でそれぞれ可  
決、成立した。

hosokawareNritsUse'ekeNga 1 meeuNo 1 ka'keta 1 seejikaikakUka'NreN 1 yoNho'oaNwa 2  
ni'juukunichi 1 go'go 3 shuusaNryo'oiN 1 hoNka'igide 2 sore'zore 1 kaketsu 3 seeritsu 1 shIta 5

関東の大手私鉄七社の年間設備投資額は売上高の五〇%前後にまで高まっており、政府は税制上  
の支援拡大が不可欠と判断した。

ka'Ntoono 1 o'ote 1 shItetsu 1 nana'shano 1 neNkaNsetsubitooshi'gakuwa 1 uriage'dakano 1  
gojuupaaseNtozeNgonimade 1 takama'tte 1 ori 1 se'efuwa 2 zeeseejoono 1 shieN 1 kakudaiga 1  
fUka'ketsUto 1 ha'NdaN 1 shIta 5

整備の主体は民間企業だが、政府にも資金面での援助措置や規制緩和による環境整備の役割を求めた。

se'ebino 1 shUtaiwa 1 miNkaNki'gyoodaga 1 se'efunimo 2 shIki'NmeNdeno 1 eNjoso'chiya 1 kIsee 1 kaNwani 1 yoru 1 kaNkyoo 1 se'ebino 1 yakuwa'rio 1 moto'meta 5

これを踏まえ、税制協の税制基本小委員会は一日、改革案の中心をなす小委報告案の骨格を作成。  
koreo 1 fuma'e 3 zeese'ekyoono 1 zeese 1 kIhoN 1 shooii'Nkaiwa 2 ichinichi 3 kaikaku'aNno 1 chuushiNo 1 na'su 1 shooihookoku'aNno 1 kokkakuo 1 sakUsee 5

大手スーパーの間で、春に一括して実施する新卒採用を見直し、採用時期を春と秋の二回に分散する動きがでてきた。

o'ote 1 su'upaano 1 aidade 2 ha'runi 1 ikkatsu 1 shIte 1 jisshi 1 suru 1 shiNsotsUsa'iyooo 1 minaoshi 3 saiyooji'kio 1 ha'ruto 1 a'kino 1 ni'kaini 1 buNsaN 1 suru 1 ugo'kiga 1 de'te 1 kIta 5

通産省・資源エネルギー庁は工場などで自家発電した電力を周辺の一般家庭や事業所向けに販売することを認める方針を固めた。

tsuusa'Nshoo 3 shigeNenerugi'ichoowa 2 koojo'onadode 2 jikaha'tsudeN 1 shIta 1 de'Nryokuo 1 shuuheNno 1 ippaNka'teeya 1 jigyooshomukeni 1 haNbai 1 suru 1 kotoo 1 mitomeru 1 hooshiNo 1 katameta 5

まず、準備段階として九三年度決算分は加入時期で格差をつけ、九四年度以降、さらに区分を拡大、九六年度にも、新方式に移行する。

ma'zu 1 juNbida'NkaitoshIte 1 kyu'ujuu 1 saNne'NdokessaNbuNwa 2 kanyuuji'kide 2 ka'kUsao 1 tsUke 3 kyu'ujuu 1 yone'Ndo 1 i'koo 3 sa'rani 1 kubuNo 1 kakudai 3 kyu'ujuu 1 rokune'Ndonimo 2 shiNho'oshIkini 1 ikoo 1 suru 5

大蔵省は、銀行と企業などの取引所を経由しない相対の金利先物取引を早ければ七月にも解禁する。

ookura'shoowa 2 giNkooto 1 ki'gyoonadono 1 torihIkijo'o 1 keeyu 1 shinai 1 sootaino 1 kiNri 1 sakimonotori'hIkio 1 ha'yakereba 1 shIchigatsu'nimo 2 kaikiN 1 suru 5

農水省は政府米の売り渡し価格（消費者米価）の決定に市場の実勢を反映させる方針を決めた。

noosu'ishoowa 2 se'efu 1 komeno 1 uriwatashIka'kaku 3 shooIshabe'eka 3 no 1 ketteeni 1 shi-joono 1 saneseo 1 haNee 1 saseru 1 hooshiNo 1 kimeta 5

se'efu 1 komeno 1 よみ誤り (seefumaino 1)  
saneseo 1 よみ誤り (jisseeo 1)

日本航空は九五年度から、国内線について勤務期間をあらかじめ限定した契約社員や派遣社員を導入する。

nihonko'okuuwa 1 kyu'ujuu 1 gone'Ndokara 1 kunina'iseNni 1 tsuite 1 kiNmukI'kaNo 1 araka-jime 1 geNtee 1 shIta 1 keeyakUsha'iNya 1 hakeNsha'iNo 1 doonyuu 1 suru 5

kunina'iseNni 1 よみ誤り (kokunaiseNni 1)

移転も考えた和泉武同事務所代表は今春になって逆に、「値下げしなければここを出る」と返答した。

iteNmo 2 kaNga'eta 1 i'zumi 1 ta'keshi 1 do'ojimusho 1 daihyoowa 1 koNshuNni 1 na'tte 1 gyakuni 1 nesage 1 shinakereba 1 kokoo 1 de'ru 3 to 1 heNto'o 1 shIta 5

藤井蔵相は地価を鎮静化するため引き上げた土地譲渡益課税の税率（現行三九%）を引き下げる考えのないことを強調した。

fujii 1 zooshoowa 1 chlka'o 1 chiNseeka 1 suru 1 tame 1 hIkia'geta 1 tochiJooteekIka'zeeno 1 zeeritsu 3 geNkoo 1 sa'Njuu 1 kyuupaase'Nto 3 o 1 hIkIsage'ru 1 kaNga'eno 1 na'i 1 kotoo 1 kyoochoo 1 shIta 5

五月調査では、先行きの予想を示す九月までの業況予測がマイナス四二と、さらに改善している。gogatsUcho'osadewa 2 sakiyukino 1 yosooo 1 shime'su 1 ku'gatsumadeno 1 gyookyooyo'sokuga 2 mainasu 1 yo'Njuu 1 ni'to 1 sa'rani 1 kaizeN 1 shIte 1 iru 5

また、安保理による経済制裁決議などの措置は北朝鮮の翻意を促すためのものであるとの認識でも一致した。

mata 3 aNpo'rini 1 yoru 1 keezaiseesaike'tsuginadono 1 so'chiwa 1 kItachoose'Nno 1 hoNio 1 unaga'su 1 tameno 1 mono'de 1 arutono 1 niNshIkide'mo 2 icchi 1 shIta 5

年間では、三十社を超え、この五年間で最高だった九〇年の三十七社とほぼ肩を並べる見通し。

neNkaNde'wa 2 saNju'sshao 1 koe 3 kono 1 gone'NkaNde 2 saikooda'tta 1 kyuuju'uneNno 1 sa'Njuu 1 nana'shato 1 ho'bo 1 ka'tao 1 naraberu 1 mitooshi 5

同時に、総裁は為替相場の円高傾向や雇用情勢など景気へのマイナス要因を引き続き注視する考えを示した。

doojini 1 soosaiwa 1 kawaseso'obano 1 eNdakake'ekooya 1 koyoojo'oseenado 1 keekie'no 1

mainasuyo'oiNo 1 hikItsuzukIchuushi 1 suru 1 kaNga'eo 1 shime'shIta 5

中国ではグループの経営コンサルタント会社、野村シティックが大企業の傘下企業の工場建設を支援する。

chu'ugokudewa 1 guru'upuno 1 keeeekoNsarutaNtoga'isha 3 nomurashI'tIkkuga 1 daiki'gyoono 1 saNkaki'gyoono 1 koojooke'Nsetsuo 1 shieN 1 suru 5

全日本空輸は米国第三位の航空会社であるデルタ航空と広範囲に業務提携することで合意、十五日基本合意書に調印した。

zeNnihoNku'uyuwa 2 beekoku 1 daisa'Nino 1 kookuuga'ishade 1 aru 1 derutako'okuuto 1 kooha'Nini 1 gyoomute'ekee 1 suru 1 kotode 2 gooi 3 ju'ugonichi 1 kIhoN 1 gooishoni 1 chooiN 1 shIta 5

トヨタ自動車は中国の有力自動車グループ「天津汽車工業公司」（天津市）と合弁で乗用車生産に乗り出す意向を固めた。

toyotajido'oshawa 1 chu'ugokuno 1 yuuryokujidooshaguru'upu 3 teNshiNkIshako'ogyoo 1 ko'oji 3 teNshi'Nshi 3 to 1 goobeNde 2 jooyooshase'esaNni 1 norida'su 1 ikoo 1 katameta 5

カーター氏が伝えた米国の意向はいずれも北朝鮮が望んでいたものだが、金主席の核問題に対する具体的な回答は伝えられていない。

kaata'ashiga 1 tsUtaeta 1 beekokuno 1 ikoowa 2 izuremo 1 kItachoose'Nga 1 nozoNde 1 ita 1 monodaga 1 kiNshU'sekino 1 kakumo'Ndaini 1 taisu'ru 1 gutaitekina 1 kaitoowa 1 tsUtaera'rete 1 inai 5

四月末から五月初めにかけて、米長期金利の急騰とドル安の連動という市場の動揺が起きた。

shigatsuma'tsUkara 1 go'gatsu 1 hajimeni 1 ka'kete 1 be'e 1 chookIki'Nrino 1 kyuutooto 1 doruyasuno 1 reNdooto 1 iu 1 shijoono 1 dooyooga 1 o'kIta 5

マツダは資本提携先の米フォード・モーターの欧州工場に、小型乗用車の生産を委託する。

matsudawa 2 shIhoNteekeesakino 1 be'e 1 foodomo'otaano 1 ooshuuko'ojooni 1 kogatajooyo'oshano 1 seesaNo 1 itaku 1 suru 5

非製造業七百七十五社の総資産が〇・〇二%と比較的小幅の減少だったのに対し、製造業千二百三十一社は一・三%減った。

hIseezo'ogyoo 1 nana'hyaku 1 nana'juu 1 go'shano 1 sooshI'saNga 2 re'eteN 1 zeroniipaase'Ntoto 1 hikakUteki 1 kohabano 1 geNshooda'ttanoni 1 ta'ishi 3 seezo'ogyoo 1 se'N 1 nihyaku 1 sa'Njuu

1 i'sshawa 2 i'tteN 1 saNpaase'Nto 1 hetta 5

ニューヨーク市場で円相場が急騰、一時一ドル＝一〇〇円を突破するなど円高や設備投資の低迷など懸念材料も残っている。

nyuuyookUshi'jooode 2 eNso'obaga 2 kyuutoo 3 ichi'ji 1 ichi'doru 1 iko'oru 1 hyaku'eNo 1 toppa 1 surunado 1 eNdakaya 1 setsubito'oshino 1 teemeena'do 1 keneNza'iryoomo 1 noko'tte 1 iru 5

これにより中国初の本格的チェーンストアが、日本最大の小売業であるダイエーを軸に実現する見通しとなった。

koreni 1 yori 1 chu'ugoku 1 hatsuno 1 hoNkakUtekIcheeNsUto'aga 2 niho'N 1 saidaino 1 kouri'gyoode 1 aru 1 daieeo 1 jiku'ni 1 jitsugeN 1 suru 1 mitooshIto 1 natta 5

NTTが申請しているのは、毎月の基本料金の平均一八%引き上げと、月二回目以降の番号案内料の上げ。

enuitiiti'iga 1 shiNsee 1 shIte 1 irunowa 2 maitsUkino 1 kIhoNryo'okiNno 1 heekiN 1 juuhachIpaaseNtohiKiageto 1 tsUki 1 nikaimo 1 i'koono 1 baNgooaNnai'ryoono 1 age 5

六月四日、ラジオ番組の収録のため大阪のFM局にいた落語家の桂文珍を一人の男が訪れた。

rokugatsu 1 yokka 3 rajioba'Ngumino 1 shuurokuno 1 tame 1 oosakano 1 efuemu'kyokuni 1 ita 1 rakugokano 1 keefumichiNo 1 hIto'rino 1 otoko'ga 1 otozu'reta 5

keefumichiNo 1                      アクセント誤り (ka'tsura 1 bunchiNo 1)

サハラ以南のアフリカ諸国など所得水準が極めて低い国の経済基盤の整備や貧困問題の解消を側面支援するのが狙い。

sa'hara 1 i'naNno 1 afurikasho'kokunado 1 shotokUsu'ijuNga 1 kiwa'mete 1 hIku'i 1 kunino 1 keezaiki'baNno 1 se'ebiya 1 hiNkoNmo'Ndaino 1 kaishooo 1 sokumeNshi'eN 1 surunoga 1 nerai 5

九五年四月の都市銀行の新規学卒者採用数は、四年連続で前年春の実績を下回ることが確実になった。

kyu'ujuu 1 go'neN 1 shigatsuno 1 toshigi'Nkoono 1 shiNkigakUsotsUshasaiyo'osuuwa 2 yoneN 1 reNzokude 2 zeNneNha'runo 1 jissekio 1 shItamawa'ru 1 kotoga 1 kakujitsuni 1 natta 5

サミット前に緊急蔵相代理会議（G7D）を開き、金利を含めたドル安防止策を検討する案も浮上している。

samittoma'eni 1 kiNkyuu 1 zooshoo 1 dairika'igi 3 jiise'buN 1 di'i 3 o 1 hira'ki 3 kiNrio 1

fUku'meta 1 doruyasu 1 booshI'sakuo 1 keNtoo 1 suru 1 a'Nmo 1 fujoo 1 shIte 1 iru 5

米政府の政策調整機関である国家経済会議（NEC）の検討結果として、ホワイトハウスなどの当局者が明らかにした。

beese'efuno 1 seesakUchooseekI'kaNde 1 aru 1 kokkakeezaika'igi 3 enuiishi'i 3 no 1  
keNtooke'kkatoshIte 1 howaitoha'usunadono 1 tookyokU'shaga 1 aki'rakani 1 shIta 5

beese'efuno 1                      アクセント誤り (be'e 1 se'efuno 1)

五月末時点での予測（一・四%低下）に比べ落ち込み幅は小さく、四一六月期は二期連続で前期比プラスになる可能性が強まった。

gogatsumatsuji'teNdeno 1 yosoku 1 ka'kko 1 i'tteN 1 yoNpaase'Nto 1 teeka 3 ni 1  
kurabeochIkomi'habawa 1 chiisa'ku 3 yo'N 1 yokoboo 1 rokugatsU'kiwa 2 ni'ki 1 reNzokude 2  
zeNklhIpu'rasuni 1 naru 1 kanooseega 1 tsuyoma'tta 5

kurabeochIkomi'habawa 1 アクセント誤り (kurabe 2 ochIkomi'habawa 1)  
yo'N 1 yokoboo 1 rokugatsU'kiwa 2  
よみ誤り (yo'N 1 kara 1 rokugatsU'kiwa 2)

九〇年からの株式相場の下落で額面割れの株式投信が相次ぎ、投信業界は九二年から償還期限を三年間延長してきた。

kyuuju'uneNkarano 1 kabushIkIso'obano 1 gerakude 2 gakumeNwareno 1 kabushIkIto'oshiNga  
1 a'itsugi 3 tooshiNgyo'okaiwa 2 kyu'ujuu 1 ni'neNkara 1 shookaNki'geNo 1 saNne'NkaN 1  
eNchoo 1 shIte 1 kIta 5

【バンコク 1日 = 竹岡倫示】 アジア諸国・地域が九四年の経済成長率見通しをこぞって上方修正している。

ka'kko 1 baNkokui'chinichi 3 takeokari'Nji 1 ka'kko 1 ajiasho'koku 3 chi'ikiga 2 kyu'ujuu 1  
yone'Nno 1 keezaiseechooritsumitooshio 1 kozo'tte 1 joohooshu'usee 1 shIte 1 iru 5

baNkokui'chinichi 3                      よみ誤り (ba'Nkoku 1 tsuita'chi 1)  
keezaiseechooritsumitooshio 1      アクセント誤り (keezaiseechooritsumi'tooshio 1)

低金利を背景に融資申し込みが急増し、住宅公庫の新規融資が難しくなっていることに伴う措置。  
teeki'Nrio 1 haikeeni 1 yuushimooshIkomiga 1 kyuuzoo 1 shi 1 juutakUko'okono 1 shiNkiyu'ushiga  
1 muzuka'shIku 1 natte 1 iru 1 koton 1 tomona'u 1 so'chi 5

yuushimooshIkomiga 1                  アクセント誤り (yu'ushimooshIkomiga 1)

住宅やオフィスの焼失・損壊を補償する火災保険の保険料引き上げが、来年二月にも実施になる  
公算が大きくなってきた。

juutakuya 1 o'fIsuno 1 shooshItsu 3 soNkaio 1 hoshoo 1 suru 1 kasaiho'keNno 1 hokeNryooohIki-  
agega 2 raineN 1 nigatsu'nimo 2 jisshini 1 naru 1 koosaNga 1 o'okIku 1 natte 1 kIta 5

円高の進展などを背景に輸入が大幅に増加したため、二半期続けての減少は、九一年度に経常  
黒字が増加に転じて以降初めて。

eNdakano 1 shiNteNna'doo 1 haikeeni 1 yunyuuga 1 oohabani 1 zooka 1 shIta 1 tamede 2  
niha'Nki 1 tsuzukete'no 1 geNshoowa 2 kyu'ujuu 1 ichine'Ndoni 1 keejooku'rojiga 1 zookani 1  
teNjite 1 i'koo 1 haji'mete 5

「地下鉄工事は当初予算を上回ることが多いが、これは我々の予想をはるかに超えている」と運  
輸省鉄道局財務課の幹部は語る。

chIkatetsUko'ojiwa 1 tooshoyo'saNo 1 uwamawa'ru 1 kotoga 1 o'oiga 1 korewa 1 warewareno 1  
yosoo 1 ha'rukani 1 koete 1 iru 3 to 1 uNyushootetsudookyokuzaimukano 1 ka'Nbuwa 1 kataru  
5

海外主幹事には野村証券の香港の現地法人、野村国際有限公司が決まり、七日上海で調印する。

ka'igai 1 shUka'Njiniwa 1 nomurasho'okeNno 1 ho'NkoNno 1 geNchIho'ojiN 3 nomura 1 kokU-  
sai 1 yuugeN 1 ko'ojiga 1 kimari 3 nanoka 1 sha'Nhaide 2 chooiN 1 suru 5

東京証券取引所は外国企業の上場基準緩和と合わせ、早ければ年内にも国内企業の子会社上場ルー  
ルを見直す。

tookyooshookeNtorihIkijo'wa 2 gaikokUki'gyoono 1 joojooki'juN 1 kaNwato 1 awa'se 1 ha'yakereba  
1 ne'Nnainimo 2 kokunaiki'gyoono 1 koga'isha 1 joojooru'uruo 1 minaosu 5

また日本企業などが発行した外債の利子の非課税措置も制度として定着していると判断、同月末  
の期限を延長する見込み。

mata 1 niho'N 1 ki'gyoonadoga 1 hakkoo 1 shIta 1 gaisaino 1 ri'shino 1 hIkazeeso'chimo 1  
se'edotoshIte 1 teechaku 1 shIte 1 iruto 1 ha'NdaN 3 doogetsuma'tsuno 1 ki'geNo 1 eNchoo 1  
suru 1 mikomi 5

建設省は九六年度から、公共工事の仕上がりぐあいに関する検査結果を公表する方針を決めた。

keNsetsU'shoowa 1 kyu'ujuu 1 rokune'Ndokara 1 kookyooko'ojiino 1 shiaga'ri 2 guaini 1 ka'Nsu'ru  
1 keNsake'kkao 1 koohyoo 1 suru 1 hooshiNo 1 kimeta 5

「不公平税制」との批判がある固定資産税や事業税などの特例措置について、九五年度税制改正で大幅な縮減を目指す。

fUkooheeze'ese 3 to'no 1 hIhaNga 2 a'ru 1 koteeshIsa'Nzeeya 1 jigyo'ozeenadono 1 tokureeso'chini 1 tsuite 1 kyu'ujuu 1 gone'Ndozeeseekaiseede 2 oohabana 1 shUkugeNo 1 meza'su 5

鉄鋼、化学、繊維、紙パなど景気に敏感な素材産業で生産が上向き、雇用削減にも歯止めがかかってきた。

tekkoo 3 ka'gaku 3 se'Ni 3 shIpana'do 1 keekini 1 biNkaNna 1 sozaisa'Ngyoode 2 seesaNga 1 uwamu'ki 3 koyoo 1 sakugeNni'mo 2 hadomega 1 kaka'tte 1 kIta 5

shIpana'do 1 よみ誤り (kamipana'do 1)

閣僚会議は、二〇二〇年の貿易自由化目標年次を打ちだしたA P E C賢人会議の報告などをもとに、貿易自由化問題を議論した。

kakuryooka'igiwa 2 nise'N 1 niyu'uneNno 1 booeeki 1 jiyuukamokUhyoone'Njio 1 uchidashI'ta 1 eepiiiishiike'NjiNkaigino 1 hookokuna'doo 1 moto'ni 1 booeeki 1 jiyuukamo'Ndaio 1 gi'roN 1 shIta 5

eepiiiishiike'NjiNkaigino 1 よみ誤り (eepe'kku 1 kenjinka'igino)

E Uに新規加盟を予定している四カ国のうち、オーストリア、フィンランドを含めた三カ国までが加盟を決めたことになる。

iiyu'uni 1 shi'Nki 1 kameeo 1 yotee 1 shIte 1 iru 1 yoNka'kokuno 1 uchi 1 oosUto'ria 3 fi'NraNdo 1 fUku'meta 1 saNka'kokumadega 1 kameeo 1 kimeta 1 kotoni 1 naru 5

【ジャカルタ 1 4 日 = 橋高聡】村山首相は十四日、ジャカルタ市内の米大使公邸で米国のクリントン大統領と一時間十五分会談した。

ka'kko 1 jakaruta 1 ju'uyokka 3 tachiba'nataka 1 sa'toshi 1 ka'kko 1 murayama 1 shUshoowa 2 ju'uyokka 3 jakarutashi'naino 1 beetaishIko'oteede 2 beekokuno 1 kuri'NtoN 1 daito'oryooto 1 ichiji'kaN 1 juugo'fuN 1 kaidaN 1 shIta 5

beetaishIko'oteede 2 アクセント誤り (be'e 1 taishIko'oteede 2)

ドル表示でも輸出の前年同月比増加率が一〇・四%だったのに対して、輸入は二四・〇%増えた。

do'ru 1 hyoojide'mo 2 yushUtsuno 1 zeNneN 1 doogetsU'hi 1 zooka'ritsuga 1 ju'tteN 1 yoNpaase'Ntodattanoni 1 ta'ishIte 1 yunyuuwa 2 ni'juu 1 yo'NteN 1 zeropaase'Nto 1 fu'eta 5

焦点の防衛費も今年度比〇・九%増の概算要求を、自衛隊発足後最低だった五七年度の伸び〇・四%以下の水準へ圧縮を目指す。

sho'oteNno 1 booe'ehimo 2 koNneNdo'hi 1 re'eteN 1 kyuupaase'Ntozoono 1 gaisaNyo'okyuuo 1 jjeetai 1 hossokugo 1 saiteda'tta 1 gojuu 1 nanane'Ndono 1 nobi 1 re'eteN 1 yoNpaase'Nto 1 i'kano 1 suijuNe 1 assuUkuo 1 meza'su 5

この政策はインドネシアを訪問中の江沢民国家主席に同行している中国政府高官が明らかにした。  
kono 1 seesakuwa 1 iNdone'shiao 1 hoomoNchuuno 1 ezawamiNkokkashU'sekini 1 dookoo 1 shIte 1 iru 1 chuugokUseefUko'okaNga 1 aki'rakani 1 shIta 5

ezawamiNkokkashU'sekini 1 よみ誤り (koutakumi'n 1 kokkashu'sekini 1)

アラファト・パレスチナ解放機構 (P L O) 議長率いる自治政府は発足から半年で最大の危機に立たされた。

arafato 1 paresU'china 1 kaihookI'koo 3 piieruo'o 3 gi'choo 1 hIkii'ru 1 jichIse'efuwa 1 hos-sokUkara 1 haNtoshi'de 2 saidaino 1 kI'kini 1 tata'sareta 5

tata'sareta 5 アクセント誤り (tatasa'reta 5)

十八日に郵政省がまとめた事業化のガイドラインに基づき、九五年中にもサービス開始に踏み切る。

juuhachinichi'ni 1 yuuse'eshooga 1 matometa 1 jigyookano 1 gaidora'iNni 1 motozu'ki 3 kyu'ujuu 1 gone'Njuunimo 2 sa'abisu 1 kaishini 1 fumiki'ru 5

gone'Njuunimo 2 よみ誤り (goneNchuuni'mo 2)

足りない電力は自家発電をしている一般企業から公開入札で購入するよう電力会社に求める。

tarinai 1 de'Nryokuwa 1 jikaha'tsudeNo 1 shIte 1 iru 1 ippanki'gyookara 1 kookainyu'usatsude 2 koonyuu 1 suru 1 yoo 1 deNryokuga'ishani 1 motome'ru 5

(小選挙区比例代表並立制は「きょうのことば」参照) = 関連記事 2、3、社会面、小選挙区立候補者の予想顔ぶれを 8、9面に。

shooseNkyo'ku 1 hiree 1 daihyoheeritsUseewa 3 kyo'ono 1 kotoba 3 saNshoo 3 kaNreNki'ji 1 ni 3 saN 3 shaka'imeN 3 shooseNkyokuritsUkoo'ho'shano 1 yosookao'bureo 1 hachi 3 kyuumeni 5

電機、自動車などすそ野の広い製造業で、大手が系列の下請け企業を選別する動きが広がってきた。

de'Nki 3 jido'oshanado 1 sUsonono 1 hiro'i 1 seezo'ogyoode 2 o'otega 1 keeretsuno 1 shI-taukeki'gyooo 1 seNbetsu 1 suru 1 ugo'kiga 1 hirogatte 1 kIta 5

ugo'kiga 1

アクセント誤り (ugoki'ga 1)

村山内閣の支持率は四二・一%、不支持は三四・六%で、ともに前回九月調査とほぼ横ばい。

murayamana'ikakuno 1 shiji'ritsuwa 2 yo'Njuu 1 ni'iteN 1 ichIpaase'Nto 3 fUshijiwa 2 sa'Njuu  
1 yo'NteN 1 rokUpaase'Ntode 2 tomoni 1 ze'Nkai 1 kugatsUcho'osato 1 ho'bo 1 yokobai 5

kugatsUcho'osato 1

アクセント誤り (kuga'tsu 1 cho'osato 1)

海外景気の拡大や猛暑の追い風に加え、リストラクチャリング（事業の再構築）の効果が出ている。

kaigaike'ekino 1 kakudaiya 1 mo'oshono 1 oikazeni 1 kuwae 3 risUtora'kUchariNgu 3 ji'gyoono  
1 saiko'ochIku 3 no 1 ko'okaga 1 de'te 1 iru 5

病気の両親などを介護するため休暇をとる介護休業制度を法制化する原案の内容が二十六日明らかになった。

byookino 1 ryo'oshiNnadoo 1 ka'igo 1 suru 1 tame 1 kyuukao 1 to'ru 1 kaigokyuugyoose'edoo  
1 hooseeka 1 suru 1 geNaNno 1 naiyooga 2 ni'juurokunichi 1 aki'rakani 1 natta 5

米国、アジアを先導役にした世界景気の拡大は出遅れていた日本、欧州を含めて、「同時好況」に向かいつつある。

beekoku 3 a'jiao 1 seNdooyakuni 1 shIta 1 se'kai 1 keekino 1 kakudaiwa 1 deoku'rete 1 ita 1  
niho'N 3 o'oshuuo 1 fUku'mete 1 doojiko'okyoo 3 ni 1 mukaitsU'tsu 1 a'ru 5

se'kai 1 keekino 1

アクセント誤り (seka'ikeekino 1)

厚年基金は企業や業界団体などが国の厚生年金に上乘せする年金給付を目的に設立する特別法人。

atsUtoshIki'kiNwa 1 ki'gyooya 1 gyookaida'Ntainadoga 1 kunino 1 kooseene'NkiNni 1 uwanose  
1 suru 1 neNkiNkyu'ufuo 1 mokUtekini 1 setsuritsu 1 suru 1 tokubetsUho'ojiN 5

atsUtoshIki'kiNwa 1

よみ誤り (kooneNki'kiN)

日米やアジアなどの企業が共同で進める大型の開発プロジェクトについて、計画づくりの段階から両国輸銀が協調、融資に応じる。

ni'chibeeya 1 a'jianadono 1 ki'gyooga 2 kyoodoode 1 sUsumeru 1 oogatano 1 kaihatsUpuro'jekUtoni  
1 tsuite 1 keekakuzu'kurino 1 daNkaikara 1 ryookokuyugiNga 2 kyoochoo 3 yu'ushini 1 oojiru  
5

ryookokuyugiNga 2

アクセント誤り (ryo'okokuyugiNga 2)

大蔵省の貿易統計で1998年の対米貿易黒字が大幅に増加したことで、日米間の貿易摩擦が再燃する懸念が高まってきた。

ookura'shoono 1 boeekIto'okeye 2 se'N 1 kyu'uhyaku 1 kyu'ujuu 1 hachi'neNno 1 taibee-boeekIku'rojiga 1 oohabani 1 zooka 1 shIta 1 kotode 2 nichibe'ekaNno 1 boeekima'satsuga 1 saineN 1 suru 1 keneNga 1 takama'tte 1 kIta 5

米議会に保護主義的な動きが高まっており、突き上げを受けた米政府も「鉄鋼摩擦」問題で強硬措置をちらつかせている。

beegi'kaini 1 hogoshugitekina 1 ugo'kiga 1 takama'tte 1 ori 1 tsUkia'geo 1 u'keta 1 beese'efumo 3 tekkooma'satsu 3 moNdaide 2 kyookooso'chio 1 chiratsUka'sete 1 iru 5

beegi'kaini 1                      アクセント誤り (be'egikaini 1)

beese'efumo 3                      アクセント誤り (be'e 1 se'efumo 3)

米国景気が失速して労働者の解雇が広がれば、次期米大統領選ともからんで、日米関係が険悪な状況に陥る可能性もある。

beekokUke'ekiga 1 shIssoku 1 shIte 1 roodo'oshano 1 ka'ikoga 1 hirohare'ba 1 ji'ki 1 be'e 1 daitooryooseNto'mo 2 kara'Nde 1 nichibeeka'Nkeega 1 keNakuna 1 jookyooni 1 ochi'iru 1 kanooseemo 1 a'ru 5

建設業界では、東急建設が若手社員を含む大規模な希望退職募集に踏み切るなど、削減計画が相次いでいる。

keNsetsugyo'okaidewa 2 tookyuu'Nsetsuga 1 wakatesha'iNo 1 fUku'mu 1 daiki'bona 1 kibootaishokubo'shuuni 1 fumiki'runado 1 sakugeNke'ekakuga 1 a'itsuide 1 iru 5

kibootaishokubo'shuuni 1                      アクセント誤り (kiboota'ishoku 1 boshuuni 1)

同時に、今年3月期決算の業績予想を下方修正、昨年11月時点で発表した経常赤字70億円の予想を250億円に変更した。

doojini 1 kotoshi 1 saNgatsUKIke'ssaNno 1 gyoosekiyo'sooo 1 kahooshu'usee 3 sakuneN 1 ju-uichigatsuji'teNde 2 happyoo 1 shIta 1 keejooa'kaji 1 nanajuuo'kueNno 1 yosooo 1 nihyaku 1 gojuuo'kueNni 1 heNkoo 1 shIta 5

また、事業ごとの売却や合併などに柔軟に対応するため、2001年度をめどに持ち株会社制に移行する方針を正式に決めた。

mata 3 jigyoogo'tono 1 baikyakuya 1 gappeena'doni 1 juunaNni 1 taioo 1 suru 1 tame 1 nise'N 1

ichine'Ndoo 1 me'doni 1 mochIkabugaishaseeni 1 ikoo 1 suru 1 hooshiNo 1 seeshIkini 1 kimeta 5

金融監督庁の銀行検査の担当者の手引書となる「金融検査マニュアル」の草案が、銀行界に波紋を呼んでいる。

kiNyuukaNtokU'choono 1 giNkooke'Nsano 1 taNto'oshano 1 tebikIshoto 1 naru 3 kiNyuukeN-sama'nyuaru 3 no 1 sooaNga 2 giNko'okaini 1 hamoNo 1 yoNde 1 iru 5

銀行が見せる困惑は、経営責任の所在をあいまいにしてきた日本企業に共通する問題を浮かび上がらせている。

giNkooga 1 mise'ru 1 koNwakuwa 2 keeeese'kiniNno 1 shozaio 1 aimaini 1 shIte 1 kIta 1 niho'N 1 ki'gyooni 1 kyootsuu 1 suru 1 moNdaio 1 ukabiagara'sete 1 iru 5

shIte 1 kIta 1                      アクセント誤り (shIte'kIta)

niho'N 1 ki'gyooni 1              アクセント誤り (nihoN 1 ki'gyooni 1)

そもそも「上司」で人事権者の代表取締役に異議を申し立てる風土は日本企業には乏しい。

so'mosomo 3 jo'oshi 3 de 1 jiNjike'Nshano 1 daihyootorishimariyakuni 1 i'gio 1 mooshItate'ru 1 fu'udowa 2 niho'N 1 ki'gyooniwa 1 toboshi'i 5

理不尽な要求だと思っても、交渉をけとばしたり、譲歩を拒み通したりできる国は、多くあるまい。

rifu'jiNna 1 yookyuuda'to 1 omo'ttemo 1 kooshooo 1 ketobashIta 1 ri 3 jo'ohoo 1 kobamito'oshItari 1 deki'ru 1 kuniwa 2 o'oku 1 arumai 5

## 13.2 good.txt

良くなった例

\*\*\* 確認用テキストファイル：数詞に関連したもの \*\*\*

\$27万4,000です。

<修正前> ni'juu nanama'N yoNseNdo'ru 1 yoNse'Ndesu 5

>修正後> ni'juu nanama'N yoNseNdo'rudesu 5

¥27万4,000です。

< ni'juu nanama'N yoNseNeN 1 yoNse'Ndesu 5

> ni'juu nanama'N yoNseNeNdesu 5

3:24です。

< sa'Nji ni'juu yo'NfuN 1 ni'juu yo'Ndesu 5

> sa'Nji ni'juu yo'NfuNdesu 5

¥0です。

< zero'eN 1 ze'rodesu 5

> zero'eNdesu 5

¥125です。

< hyaku ni'juu go'eN 1 hyaku ni'juu go'desu 5

> hyaku ni'juu go'eNdesu 5

\$1です。

< ichido'ru 1 ichi'desu 5

> ichido'rudesu 5

\$1です。

< ichido'ru 1 ichi 3 de'su 5

> ichido'ru 3 de'su 5

\$100です。

< hyakudo'ru 1 hyaku'desu 5

> hyakudo'rudesu 5

\*\*\* 確認用テキストファイル：数詞の多く出てくるもの \*\*\*

生鮮食品を除いたベースで101.6となり、

< seeseN 3 shokUhiNo nozoita be'esude 2 hyaku i'tteN roku 1 rokU'to nari 1  
.....

> seeseN 3 shokUhiNo nozoita be'esude 2 hyaku i'tteN rokU'to nari 1  
.....

〒430-8650

< yuubiNba'Ngoo yoNsa'N ze'rono hachiro'ku gooze'ro 1 hasse'N roppyaku goju'u 3  
.....

> yuubiNba'Ngoo yoNsa'N ze'rono hachiro'ku gooze'ro 3  
.....

03-3292-2934

< zerosa'Nno saNni'i kyuuni'ino niikyu'u saNyo'N 1 nise'N kyu'uhyaku sa'Njuu yo'N 3  
.....

> zerosa'Nno saNni'i kyuuni'ino niikyu'u saNyo'N 3  
.....

TEL 045-451-2326

< zeroyo'N go'no yoNgo'o ichi'no niisa'N niiro'ku 1 nise'N sa'Nbyaku ni'juu roku 3  
.....

> zeroyo'N go'no yoNgo'o ichi'no niisa'N niiro'ku 3  
.....

TEL 03-5702-9136

< zerosa'Nno goona'na zeroni'ino kyuui'chi saNro'ku 1 kyuuse'N hyaku sa'Njuu roku 3  
.....

> zerosa'Nno goona'na zeroni'ino kyuui'chi saNro'ku 3  
.....

\*\*\* 確認用テキストファイル：一般的なもの \*\*\*

数千から数万人規模の反米英デモが続いている。

.....  
(修正前) suuse'N 1 seNka'ra 1 su'u 1 maNniNki'bono haNbeeeede'moga tsuzuite iru 5  
.....

(修正後) suuse'Nkara 1 su'u 1 maNniNki'bono haNbeeeede'moga tsuzuite iru 5  
.....

筒井昭宣教頭は「万一を考えれば、  
.....

(修正前) tsUtsui akiraseNkyooa'tamawa 3 ichi 1 ichi'o kaNgae'reba 1

~~~~~

(修正後) tsUtsui akiraseNkyooa'tamawa 3 ichi'o kaNgae'reba 1

~~~~~

\* 「万一 (まんいち)」が登録されておらず、「0万1 (=いち)」と解釈され、「イチ」という読みが付けられてしまっている。

また、形態素解析に失敗している部分がある

「(誤) 筒井昭 (つついあきら) 宣教 (せんきょう) 頭 (あたま)」

「(正) 筒井昭宣 (つついあきのり) 教頭 (きょうとう)」

### 13.3 CVS 履歴

次頁より、本報告書に関連した CVS 履歴を添付する。

それぞれ、morph/ChangeLog および gfilter/ChangeLog からの抜粋である。

なお、修正前のソースコードには、CVS タグ stable\_990114 が付けられている。修正前のソースコードを参照する必要がある場合、

```
cvs checkout -r stable_990114 morph      (mor2bou)
```

```
cvs checkout -r stable_990114 gfilter   (gfilter)
```

により取り出しが可能である。

## mor2bou の CVS 履歴 (抜粋)

Date: 1999.01.14

Module: mor2bou

対応者: Sado

修正: /ou/->/oo/, /ei/->/ee/ の修正

修正ファイル: PROSODY/cal\_accbou/conv.c[1.5]

新規ファイル: PROSODY/cal\_accbou/kana\_def.h[1.1]

Date: 1999.01.25

Module: mor2bou

対応者: Sado

修正: Perl 版無声化フィルタの C への移植と mor2bou への組み込み

修正ファイル: PROSODY/cal\_accbou/hfr.c[1.2]

PROSODY/cal\_accbou/makefile[1.5]

新規ファイル: PROSODY/cal\_accbou/unvoice.c[1.1]

Date: 1999.01.25

Module: mor2bou

対応者: Sado

修正: mor2bou にスイッチ '-U' (無声化しない) を追加

修正ファイル: PROSODY/cal\_accbou/hfr.c[1.3]

PROSODY/cal\_accbou/mor2bou[1.5]

Date: 1999.01.25

Module: mor2bou

対応者: Sado

修正: 文字コードがおかしかったので修正

「中米向け援助」のアクセントの修正

(/chuubeemuke'eNjo/ -> /chuubeemukee'Njo/)

修正ファイル: PROSODY/cal\_accbou/calaccent.c[1.7]

Date: 1999.02.01

Module: mor2bou

対応者: Sado

修正: 「\$1 です。」を「いちどるいちです」と読むバグを修正。

ただし、不完全で、ドルの後に、BI 区切り (1) が入る。

修正ファイル： PROSODY/cal\_accbou/num\_kana.c[1.6]

Date: 1999.02.01

Module: mor2bou

対応者： Sado

修正： 上記バグ修正時にソースコードに挿入した emacs 用のタグ幅指示の  
コマンド記述のミスを修正

修正ファイル： PROSODY/cal\_accbou/num\_kana.c[1.7]

Date: 1999.02.01

Module: mor2bou

対応者： Fujisawa

修正： version.c を追加

新規ファイル： PROSODY/cal\_accbou/version.c[1.1]

修正ファイル： PROSODY/cal\_accbou/makefile[1.6]  
PROSODY/cal\_accbou/mor2bou.c[1.6]  
PROSODY/README[1.8]

Date: 1999.02.02

Module: mor2bou

対応者： Fujisawa

修正： decl でしか BI5 を入れていなかったため、文末の  
「?」などで BI5 を入れていなかった。

decl/ques/excl(文境界) で BI5 を入れるように修正

修正ファイル： PROSODY/cal\_accbou/version.c[1.2]  
PROSODY/cal\_accbou/conv.c[1.6]

Date: 1999.02.02

Module: dictionary patch

対応者： Sado

修正： 「1羽、2羽」を「イッパ、ニハ」と読むので、形態素解析辞書より、  
数詞の「羽(は)」を取り除き、「羽(わ)」のみにした。

「イチワ、ニワ、サンバ」と読む。

修正ファイル： KANJI/base.dic[1.8]

Date: 1999.02.04

Module: mor2bou

対応者： Sado

修正： 「\$1です。」を「いちどるいちです」と読むバグを再修正。

「3:25です」も「3じ25ふん25です」と読まなくなった。

修正ファイル： PROSODY/cal\_accbou/num\_kana.c[1.8]

Date: 1999.02.04

Module: dictionary patch

対応者： Sado

修正： 「万ー」を「イチ」と読むので、形態素解析辞書を修正。

「万ー (まんいち)」 510 -> 500(base.dic), 100(add.dic)

また、「昭宣(あきのり, 人名)」を add.dic に追加。

修正ファイル： KANJI/base.dic[1.9]

KANJI/add.dic[1.13]

Date: 1999.02.05

Module: mor2bou

対応者： Sado

修正： 文末の「です/ます」を無声化しないバグがあったので、修正。

修正ファイル： PROSODY/cal\_accbou/unvoice.c[1.2]

Date: 1999.02.05

Module: mor2bou

対応者： Sado

修正： ドルのアクセント位置の修正（「いちど'る」 -> 「いち'どる」）

修正ファイル： PROSODY/cal\_accbou/num\_kana.c[1.9]

Date: 1999.02.08

Module: mor2bou

対応者： Sado

修正： 今まで BI が存在しなかった区切りに、BI(1) を出力するよう変更。

修正ファイル： PROSODY/cal\_accbou/mor2bou.c[1.6]

Date: 1999.02.08

Module: morph

対応者： Fujisawa

修正： rcsid を ChangeLog に追加, morph に組み込み

修正ファイル： PROSODY/KANJI/makefile[1.5]

Date: 1999.02.08

Module: mor2bou

対応者: Fujisawa

修正: rcsid を ChangeLog に追加, mor2bou に組み込み  
mor2bou の version.c を削除

削除ファイル: PROSODY/cal\_accbou/version.c

修正ファイル: PROSODY/cal\_accbou/makefile[1.7][1.8]  
README[1.9]

Date: 1999.02.08

Module: morph

対応者: Fujisawa

修正: 以下の foo を morph に通すと core dump する.  
KANJI/makefile で /bin/cc を使うようにコメント  
を追加.

/bin/cc でコンパイルすると OK, gcc-2.7.2.2 は NG

```
$ cat foo
```

あらゆる現実をすべて自分のほうへねじ曲げたのだ。

一週間ばかりニューヨーク取材した。

```
$ kan2rom < foo
```

arayu'ru 1 geNjitsuo 1 su'bete 1 jibuNno 1 hooe 1 nejima'getanoda 5

```
$ ls core
```

core

修正ファイル: KANJI/makefile[1.6]

Date: 1999.02.09

Module: mor2bou

対応者: Sado

修正: BI が多く入りすぎる ([1.6] でエンバグ) 場合があるので修正。

修正ファイル: PROSODY/cal\_accbou/mor2bou.c[1.7]

Date: 1999.02.12

Module: mor2bou

対応者: Sado

修正: 「1ドル」(カタカナ表記)のアクセントを修正。

num\_kana.c[1.9]の修正では、記号「\$」の時だけでした。

修正ファイル： PROSODY/cal\_accbou/calaccent.c[1.8]  
PROSODY/cal\_accbou/inritu.h[1.4]

Date: 1999.02.12

Module: mor2bou

対応者： Sado

修正： 「1\$ です」を「いちです」と読むバグを修正。

修正ファイル： PROSODY/cal\_accbou/num\_kana[1.10]

Date: 1999.02.15

Module: mor2bou

対応者： Sado

修正： 「%」に「パーセント」と同じアクセントが付くよう修正。

「パーソナル・コンピューター」も修正。

修正ファイル： PROSODY/cal\_accbou/calaccent.c[1.9]

PROSODY/cal\_accbou/inritu.h[1.5]

Date: 1999.02.15

Module: dictionary patch

対応者： Fujisawa

修正： 危険なパッチなので cstar.patch で対応

TAS80221 「外国の方にも人気があります。」

-> gaikokuno ho'o

「外国の方」(がいこくのかた)を cstar.patch に追加

TZS80032 「天気予報では二三日は晴れていると言っていました。」

-> ni'juusaNnichiwa

「二三日」(にさんにち)を追加

「二三日」(二十三日)の優先度を下げた

修正ファイル： KANJI/README.cstar[1.2]

KANJI/cstar.patch[1.3]

Date: 1999.02.17

Module: mor2bou

対応者： Sakai (Commit は Fujisawa)

修正： 付属語検出エラーからのリカバリ機能をいれてコンパイル

すると、サ変名詞+「する」でエラーになるバグを修正。

これにともない、-DIGNORE\_FUZOKUGO\_ERROR でコンパイル

したときの不具合が無くなったため、-DIGNORE\_FUZOKUGO\_ERROR を  
廃止。README から gfilter 用にコンパイルするときの  
記述を削除。

修正ファイル： PROSODY/cal\_accbou/inritu.c[1.8]  
                  PROSODY/srh\_ftbl/srh\_ftbl.c[1.8]  
                  PROSODY/makefile.option[1.7]  
                  README[1.10]

Date: 1999.02.17

Module: mor2bou

対応者： Sado

修正： 1000 ドルが /seNdoru/ となっていたのを /se'Ndoru/ と修正。  
      また、10000000 円が「せんまんえん」は変なので、  
      「1000」は /se'N/、 「10000000」以上の時は /isse'N/ とするようになった。

修正ファイル： PROSODY/cal\_accbou/number\_accent.c[1.4]

Date: 1999.02.22

Module: mor2bou

対応者： Sado

修正： -DMATRIX でコンパイルしたときの不具合を修正。  
      「四千三百二十一です」を「よんせんさんびやくにじゅういちです」  
      「四三二一です」を「よんさんにいいちです」と読むようにした。

修正ファイル： PROSODY/cal\_accbou/num\_kana.c[1.11]

Date: 1999.02.23

Module: mor2bou

対応者： Sado

修正： gfilter から呼ばれたときに unvoice が正常に動作するよう修正

修正ファイル： PROSODY/cal\_accbou/unvoice.c[1.3]

## gfilter の CVS 履歴 (抜粋)

Date: 1999.2.23 xssado

CHATR に合わせ BI を常に挿入するよう変更

src/tdmt\_mor2bou.c [1.3]

Date: 1999.2.24 xssado

無声化処理を Perl 版呼び出しから、C 版の unvoice() を呼び出すよう変更

src/makefile [1.4]

src/gfilter.c [1.5]

## 13.4 数詞処理

「韻律処理マニュアル」に基づき、数詞・助数詞の処理を記述する。  
数字相当語句に関するアクセント結合の説明は省略した。

### 13.4.1 流れ

mor2bou では、

1. 数字への読み・アクセントの付与
2. 数字+単位数字<sup>1</sup>へのアクセントの付与

を行っている。数字・単位数字が続いたものを数詞と呼ぶ。数詞には、「桁読み」と「分離読み」があり、条件に基づき読み分けられている。

定められた品詞 (13.4.4節 参照) が数詞に後続した場合、

1. 数詞+助数詞へのアクセントの付与
2. 数詞+助数詞の音便処理

が行われる。これを助数詞処理と呼ぶ。

---

<sup>1</sup> 「万」、「億」などの桁を表す漢数字を、単位数字と呼ぶこととする

### 13.4.2 数詞

数字・単位数字が続いたものを数詞と呼ぶことにする。本節では、数字のアクセント型を定義する。

#### 1桁の数字

1桁の数詞は、次の読み・アクセント型を持つものとする(表13.1)。

表 13.1: 数字1桁の読み・アクセント

数字	桁読み	分離読み	アクセント型
0	ゼロ	—	1
1	イチ	—	2
2	ニ	ニイ	1
3	サン	—	0
4	ヨン	—	1
5	ゴ	ゴウ	1
6	ロク	—	2
7	ナナ	—	1
8	ハチ	—	2
9	キュウ	—	1

#### 2桁以上の数字

2桁以上の数字は次の読み・アクセント型を持つものとする(表13.2)。ここに示す以外の数字の組み合わせは、複数のアクセント句に分割される。

表 13.2: 数字 2 桁以上の読み・アクセント

数字	読み	アクセント型	数字	読み	アクセント型
11	ジュウイチ	4	10	ジュウ	1
12	ジュウニ	3	20	ニジュウ	1
13	ジュウサン	1	30	サンジュウ	1
14	ジュウヨン	1	40	ヨンジュウ	1
15	ジュウゴ	1	50	ゴジュウ	2†
16	ジュウロク	4	60	ロクジュウ	2†
17	ジュウナナ	3	70	ナナジュウ	2
18	ジュウハチ	4	80	ハチジュウ	2†
19	ジュウキュウ	1	90	キュウジュウ	1

数字	読み	アクセント型	数字	読み	アクセント型
100	ヒャク	2†	1000	セン	1‡
200	ニヒャク	3†	2000	ニセン	2
300	サンビャク	1	3000	サンゼン	3
400	ヨンヒャク	1	4000	ヨンセン	3
500	ゴヒャク	3†	5000	ゴセン	3
600	ロツピャク	4†	6000	ロクセン	2
700	ナナヒャク	2	7000	ナナセン	3
800	ハツピャク	4†	8000	ハツセン	3
900	キュウヒャク	1	9000	キュウセン	3

†… 後続する数字があるときは 0 型

‡… 読みが「イッセン」の時は 3 型

### 13.4.3 数字と単位数字のアクセント結合

四種類の漢数字「万」、「億」、「兆」、「京」を単位数字と呼ぶ。これら単位数字が数字に後続した場合のアクセント型を次に定義する(表 13.3, 表 13.4)。

表 13.3: 桁読みの数字と単位数字のアクセント結合

先行数字	万	億	兆	京
1, 2, 3, 4, 5, 6, 7, 8, 9	N1 + 1	N1	N1	N1
10	N1 + 1	N1	N1	N1
11, 12, 13, ..., 19	N1 + 1	N1	N1	N1
20, 30, 40, ..., 90	N1 + 1	N1	N1	N1
200, 300, 400, ..., 900	N1 + 1	N1	N1	N1
1000	N1 + 1	N1	0 <sup>†</sup>	0 <sup>†</sup>
2000, 3000, 4000, ..., 9000	N1 + 1	N1	0	0

†... 読みが「イッセン」の時は N1 型

N1 は、先行数字のアクセント型

表 13.4: 分離読みの数字と単位数字のアクセント結合

万	億	兆	京
N1 + 1	N1	N1	N1

#### 13.4.4 助数詞処理

名詞・接尾語または助数詞が数詞に後続した場合、数詞と後続単語をアクセント結合し、音便の処理を行う。これを助数詞処理と呼ぶ。助数詞処理について定義する。

##### 助数詞処理の対象となる品詞

数字と数字のアクセント結合、および、数字と、それに後続する名詞・接尾語・助数詞のアクセント結合は、助数詞処理として扱われる。

ただし、名詞のうち

- 固有名詞
- 非固有名詞で、形容動詞的名詞または副詞的名詞

は、助数詞処理の対象とならない。<sup>2</sup>

##### 助数詞とアクセント結合するアクセント句

数字処理で作成されたアクセント句のうち、助数詞に最も近いもの一つだけが、助数詞とアクセント結合する(13.5)。

表 13.5: 助数詞のアクセント結合対象となる数詞

312 個	サ'ンビャク <u>ジュウニ'コ</u>
322 個	サ'ンビャク ニ'ジュウ <u>ニ'コ</u>
312 万個	サ'ンビャク <u>ジュウニマ'ンコ</u>
322 万個	サ'ンビャク ニ'ジュウ <u>ニマ'ンコ</u>

(下線部がアクセント結合の対象)

##### 助数詞の分類

品詞「助数詞」は、数字とのアクセント結合の種類によって分類されている(表 13.7)。<sup>3</sup>

<sup>2</sup>理由は不明

<sup>3</sup>「数字処理の流れ」による定義と辞書の記述が異なる場合がある。どちらかの誤りかも知れないが、多くは、どちらも誤りとはいえない場合だと思われる。韻律処理マニュアル p.15 を参照。

この分類は辞書に記載された品詞情報によって与えられ、具体的にはアクセント接続情報下位1バイト(図13.6)によって指示される(韻律処理マニュアル, p.16)。助数詞を伴う数詞に読みとアクセントを与える関数 num\_kana() は、この種類をパラメータとして呼び出される(韻律処理マニュアル, func-64)。

表 13.6: 辞書に記載された品詞情報 (base.dic より抜粋)

読み	表記	品詞情報
ドル	ドル	910000100000 <u>0</u> 210:>>N
サツ	冊	910000100001 <u>1</u> 020:>>N
キャク	脚	910000100001 <u>1</u> 020:>>N
キョク	曲	910000100001 <u>1</u> 020:>>N
カイ	回	910000100001 <u>3</u> 020:>>N
テン	点	910000100001 <u>4</u> 010:>>N
テン	点	910000100001 <u>4</u> 010:>>N
ダイ	台	910000100000 <u>2</u> 010:~::~=N
ホン	本	910000100000 <u>2</u> 020:>>N
ダイ	台	910000100000 <u>2</u> 010:~::~=N
ネン	年	910000100000 <u>5</u> 010:D:>>N:>>R:=G
エン	円	910000100000 <u>7</u> 020:>>N
ガタ	型	910021100000 <u>9</u> 001:>>N:>>
カイ	階	910000100000 <u>6</u> 020:>>N
ショウ	勝	910000100000 <u>6</u> 020:=:>>N

↑  
アクセント接続情報下位1バイト

表 13.7: 助数詞の分類表

助数詞の種類	(*)	助数詞の例
一般	0	ドル ヶ条 ヶ月 トン 羽(ワ) 貫 間(ケン?) 気斤(キン) 筋(すじ) 句 区間 軒 個 校 差し 才 歳 皿 市 死 紙 手 寸 選 銭 層 通 斗 灯 頭 波 匂
特殊 A	1	冊 色 隻 足(ソク) 箔 匹 疋
特殊 A'	b	脚 曲(キョク) 室(シツ) 尺(シャク) 石(コク) 節(セツ) 着(チャク) 発(ハツ)
特殊 B	2	回 点
特殊 C	3	台 本 枚
特殊 D	4	年
特殊 D'	c	人
特殊 E	5	円
特殊 F	6	型(ガタ) 重(ジュウ) 大(ダイ) 丁目 年後 年産 倍分(ブン)
特殊 F'	9	階 級 週 面
特殊 F''	d	勝 審 戦(セン) 等(トウ) 敗(ハイ)

(\*) ... アクセント接続情報下位1バイト

## 数詞と後続単語のアクセント結合

助数詞処理で後続単語が数詞とアクセント結合する場合、後続単語のアクセント型などにより、数詞とのアクセント結合の種類は次の4つに分類される。

1. 助数詞が、漢字1字、もしくは、カタカナ・ひらがな2モーラ以下の場合  
ただし、カタカナ・ひらがな2モーラ以下の場合、次の条件

- 長音・撥音により2モーラとなっている
- 辞書に記された特定単語(表13.9)<sup>4</sup>

のいずれかを満たさなければならない

2. 条件1.に該当せず、助数詞のアクセント型が、平板型、もしくは、頭高型
3. 条件1.に該当せず、助数詞のアクセント型が、中高型

と分類される。

助数詞が条件1.に該当した場合、アクセント結合の型はさらに2種類に分類される。

- 先行数字は、単位数字でない(1a.)
- 先行数字が「万」、「億」等の単位数字である(1b.)

後続単語ははこれら4種類(1a, 1b, 2, 3)に分類され、数字とアクセント結合される。(表13.8)

---

<sup>4</sup>ここで、「辞書に記された特定単語」であるかどうかの判定は、KDD mor2bouでは、韻律処理構造体配列中、助数詞のアクセント接続情報上位1バイトが、2である、という条件を満たす場合、「辞書に記された特定単語」としている。これは、助数詞が2モーラのカタカナで、数詞に低くつく性質を持つとき、2にされる。

表 13.8: 助数詞処理による数字と後続単語のアクセント結合

先行数字	1a	1b	2	3
0, 1, 2, 3, 4, 5, 6, 7, 8, 9	N1	0	N1-1	N1+M2
10	N1	0	N1-1	N1+M2
11, 12, 13, ..., 19	N1	N1-1	N1-1	N1+M2
20, 30, 40, ..., 90	N1	N1-1	N1-1	N1+M2
100	N1	0	N1-1	N1+M2
200, 300, 400, ..., 900	N1	0	N1-1	N1+M2
1000	N1†	N1 - 1†	N1-1	N1+M2
2000, 3000, 4000, ..., 9000	0	N1-1	N1-1	N1+M2

N1 は先行数字のアクセント型、M2 は後続助数詞のモーラ数

†... 読みを「セン」と仮定したが、「イッセン」の場合 0 型

‡... 読みを「イッセン」と仮定したが、「セン」の場合 N1-1 型

表 13.9: 辞書に記された特定単語 (部分, base.dic より抜粋)

読み	表記	品詞情報
ドル	ドル	91000010000000210:>>N
ミリ	ミリ	10000010000000210:>>N
キロ	キロ	1000001000000a210:>>N:˘
		↑
		アクセント接続情報上位 1 バイト

### 13.4.5 数字と特殊助数詞のアクセント結合

品詞「助動詞」である単語のうち、表 13.7に該当するものを、特殊助数詞とする。助数詞処理による数詞と後続単語のアクセント結合において、後続単語が特殊助数詞に該当する場合、数詞とのアクセント結合の規則が変更される。特殊助数詞が数詞に後続した場合のアクセント結合の規則を表 13.10に示す。

表 13.10: 数字と特殊助数詞が結合した場合のアクセント結合規則の変化

0	一般	A	A'	B	C	D	D'	E	F	F'	F''
0	—	—	—	—	—	—	—	—	0	0	0
1	—	N1+N2	N1+N2	—	—	—	—	0	0	0	0
2	—	—	—	N1+N2	—	—	—	—	0	0	0
3	—	—	—	N1+N2	—	0	N1+N2	0	0	0	0
4	—	—	—	N1+N2	—	0	N1+N2	—	0	0	0
5	—	—	—	N1+N2	—	0	N1+N2	—	0	0	0
6	—	N1+N2	N1+N2	N1+N2	—	—	—	—	0	0	0
7	—	—	—	N1+N2	—	—	—	—	0	0	0
8	—	N1+N2	N1+N2	N1+N2	—	—	—	—	0	0	0
9	—	—	—	N1+N2	—	—	—	—	0	0	0
10	—	N1+N2	N1+N2	N1+N2	—	—	—	0	0	0	0
11	—	N1+N2	N1	—	—	—	—	0	0	0	N1
12,13,14, ..., 19	—	—	—	—	—	—	—	—	0	0	N1
20,30,40, ..., 90	—	—	—	—	—	—	—	—	0	N1	N1
100	—	N1+N2	N1+N2	N1+N2	—	—	—	—	0	0	0
(200 以上)	—	—	—	—	—	—	—	—	—	—	0

### 13.4.6 音便処理

韻律処理マニュアル付録部分 p.32 の「表 助数詞音韻変化表」を参照。音韻変化の規則として、品詞が助数詞の場合「(1) 助数詞」が適用され、品詞が名詞・接尾辞の場合「(2) 一般ルール」が適用される。