

日本語の文節の認定

坂 本 義 行
(電子技術総合研究所)

1. はじめに

計算機を利用する日本語のテキスト処理では、その操作基準となる語彙単位を決定することが不可欠である。現在では、前処理として、人間の手を借りて、何らかの単位に分割するといった計算機の特性を半減させるような、ぼう大な時間・労力、費用を要する方法が採られている。

一方、これを計算機で自動的に行う方法として、「分かち書き」¹⁾「自動単位分割」^{2),3)}「単語の認定」⁴⁾その他^{5),6),7)}の研究が発表されている。これらの研究は、その目的や応用によって、各々異なる方法が採られている。また、言語研究の一貫としての実験的なものから比較的実用化が容易なものまである。

その方法は、以下の2つに大別される。

i) ぼう大な辞書あるいはテーブル類を用いて検索を行う。

ii) 簡単な構文規則や語結合法則をプログラム中に組み込み処理を行う。

i) の方法は、自立語を中心にすべての語彙を辞書に登録するため、辞書の作成における質と量が問題となる。しかしプログラム自体は、辞書検索が中心で、汎用的に用いることができる。処理効率の点からは、ii) に及ばない。現在までに開発されたものは、研究あるいは実験的なシステムである。

ii) の方法は、付属語を中心に、辞書、テーブル類をできるだけ用いず、処理規則はプログラム中に機能的に組み込まれている。このため処理効率は高いが、規則の修正はプログラムの変更を要し、質を高めるには、プログラムが複雑となり、汎用的でない。しかし粗い精度の実用的なシステム向きといえる。

いずれの方法においても、最適な分割を行うには、そのテキストの形態、構文、意味に関わる情報を与えて、テキスト全体にわたる解析を方向づける必要がある。また、その解析の手順として、階層的で、各々の処理ルーチンは独立に動作すべきである。

こゝでは、その第1段階の処理として、i) と ii) の折衷的な方法として、

iii) 字種による識別と小規模な辞書あるいはパターン表を用いる分割を基本とした「文節」の認定

といった実用化を目標としたシステムを開発したので、こゝで報告いたします。

2. 方 法

これまでに開発されてきたいくつかの方法ならびにその特徴につけては、文献⁸⁾で比較、検討を行なっているので、こゝでは省略いたします。たゞ、その方法は、利用目的によつて異なってくるとおもわれるが、試みられた方法は、ほとんど実験的なもので、多量のテキストにつけて定量的な結果が示されていなかった。

私が採った方法は、前述の ii) の方法に近く、精度の向上(99%以上)、実用化を目標としてシステム開発を行なった。処理対象とするテキストの分野を限定することにより、語彙及び表記の特徴を知る。とくに平仮名連系と自立語の漢字仮名混じり表記について調査を行なった。

テキストを科学技術文献、さらに分野を限定し、今回の実験では、特許公報、

第12類、金属の加工に関する約40件(約30万字)を用いた。このテキストについて、ALPS-CONCORDANCEシステム⁹⁾を用いてKWICを作成し、約9万行を漢字アプリエツで出力した。その結果を用いて、次節に述べるような実体調査を行なった。得られた調査結果を分析し、必要な情報を辞書、テーブル類の資料とした。

処理方法として、文字種(平仮名連系から非平仮名連系への変り目)による基本分割を行なう。この区切り目の前後の結合関係についての情報を辞書及びテーブルの形で与え、修正処理を施す。分割された単位は文節(α単位¹⁰⁾)に近いものである。文節程度の大きな単位を採用したのは、この段階での処理は、次の処理が可能な範囲で最長のものにする。これにより応用範囲の拡大と単位切りされたものが切れすぎないようにする。目的によって語彙の詳細な分析を必要とする場合には、自立語・付属語辞書、構文分析、意味分析を行なって、精度を上げるといった目標を設定した。

3. 実体調査

3.1 字種見出しの割合

標本として、特許公報1冊、本文の42万5千文字について、漢字仮名混じり文を入力とし、字種分割によるKWICを作成した。この結果をもとに人手により、各字種間の接続数および各字種見出し(字種の変り目から次の字種の変り目までの連系)を調べた結果、第1表に示すような値が得られた。

なお、字種見出しの割合を右欄かつこ内に示した。この割合と比較すべき同一の例は他に見出せず、新聞についての野村氏の字種による割合を示した。この例は、字種による文字の出現率であり、われわれの調査した見出し数とは単純比較はできない。この表から、技術文献の特徴として、句読点、英字、数字が新聞の場合に比較して多い。また、文字種、字種見出しでの漢字と平仮名との比がほぼ1対1、平仮名のほうがやや多いという興味ある結果が得られた。

3.2 分割の精度

字種の変り目で分割した場合に、どの程度の誤りが発生し、その内訳はどのようなものとなるかを調べた。異なる字種間の結合関係のうち、平仮名列から非平仮名列への変り目で分割することを前提とした。

3.2.1 調査のポイント

1) 切れるべきところとそうで切れていない場合

第1表 各字種見出し数及び接続数

	句読点	漢字	平仮名	数字	平仮名	英字	ギリシア	ローマ	特殊記号	間隔	計	割合 (字種割合)
句読点	0	2838	691	205	1169	188	1	0	45	1282	6419	2.6
漢字	828	0	1352	2001	26467	157	1	0	621	341	31768	2.3
平仮名	321	2016	0	617	1837	69	0	0	177	36	5073	2.5
数字	38	1082	263	0	1331	435	22	0	2308	320	5799	2.4
平仮名	5038	23628	2385	915	0	373	6	0	170	23	32538	2.3
英字	45	211	106	215	488	0	0	0	927	60	2052	2.3
ギリシア	2	7	6	0	7	8	0	0	9	0	39	0.04
ローマ	0	0	0	0	0	0	0	0	0	0	0	0
特殊記号	148	750	181	1506	970	776	9	0	45	4385	4385	2.7
間隔	0	1236	89	339	269	46	0	0	128	0	2107	2.3
計	6420	31768	5073	5798	32538	2052	39	0	4385	2107	90180	

2) 切れるべきでないところでは切れている場合

3) 全体として字種による文節分かち書きを行なった場合の分割精度

3.2.2 調査結果

i) 切れていない場合

1) 同一字種内

a. 漢字列*

$$700 / 31,768 = 2.2\%$$

b. 平仮名列

$$2,746 / 32,538 = 8.4\%$$

ii) 字種の変わり目

a. 漢字, 片仮名, 数字, 英字相互間

$$133 / 8,524 = 1.6\%$$

a1. 漢字副詞, 英字助数詞が先行する場合

$$96$$

a2. 漢字接統詞が後続する場合

$$37$$

b. 漢字, 片仮名, 数字, 英字から平仮名

$$553 / 30,123 = 1.8\%$$

b1. 漢字副詞等が平仮名自立語に先行する場合

$$52$$

b2. 平仮名接統詞が後続する場合

$$501$$

2) 切れるべきでない場合

$$1,551 / 26,723 = 5.8\%$$

3.2.3 分割精度**

1) 切れるべきところで切れない箇所

$$4,132$$

2) 切れるべきでないところでは切れている箇所

$$1,551$$

3) 句読点を含め, 切れるべき箇所の数(文節数)

$$35,723$$

$$4) \text{ 分割の精度} = \left\{ 1 - \frac{4132 + 1551}{35723} \right\} \times 100 = 84\%$$

この分割精度が高いか低いかには異論があろう。しかし実用という面から考えれば、16%の誤りが発生するということは、1頁2,000字として、その文節数を340とすれば、54個もの誤りが発生することになる。人手による後処理を考えると1% (3.4) 以内の誤りでないと実用には供し得ないと考えられる。すなわち、実用という目的から修正処理によって、99%以上の精度が得られるかについて実験を行なった。

なお、分かち書きの基準として問題となった点について、以下のように定めた。

1) 形式名詞は付属語とある

形式名詞と実質名詞との境は、必ずしも判然としないが、一応の基準を定めた。また、平仮名表記と漢字表記ともに認める。

2) 普通名詞+数詞

「図1」、「実施例5」、「モノエタノールアミン60cc/l」、「温度30℃」など、数量を表わすものも、前の語と合わせて1文節とする。

3) 形式用言は狭い範囲に限定する。

「しなければならぬ」、「し得る」のように助動詞に接尾辞としてほゞ定着したもの、「てりる」、「て居る」、「ておく」のように学識的に補助用言とみなされるもののみに限る。余り実質的な意味をもたない用言も文節とみなす。

例：と\いう、れば\よい、ことに\なる

*連続した漢字連系の両端のみについて調べ、漢字連系全体については調べない。

** 特殊記号の前後における分割の判定は複雑であり、個々の場合について検討する必要がある。調査の時点でその判定方法が決定していなかったために、この統計値に含まれていない。

4) 特殊記号について

- i) 区切り符号は句読点と同じ 例: 「、」、「。」、「…」、「空白」、「@」
- ii) かっこ類, 演算子, 結合子, 助数詞, 添字, ギリシア字, ローマ数字, 単位は漢字種と同じ。
- iii) 多用記号類は, その前後の文字連系により決定する。
例: 「-」、「。」、「,」、「。」

4. 実験

4.1 修正情報

調査で得られた結果をもとに以下のよう修正処理のための表および辞書を作成した。

1) 見出し表 (付表1)

- a) 漢字副詞表 例: 且, 又, 充分, 従来 ----- 58
- b) 名詞接続詞表 例: 及び, 並びに, 又は ----- 9
- c) 助数詞 例: %以上, 分間, 回転 ----- 20
- d) 特殊記号 例: …, (, °C ----- 59

2) 辞書 ----- 696 (見出し)

平仮名, 漢字, 漢字仮名混じりのいかなるものを問わず, 付属語に属するものは, なるべくもうらう。即ち, 活用語尾, 助詞, 助動詞, 動詞性又は形容詞性接辞, 補助用言, 形式名詞等である。その他, 平仮名又は漢字仮名混じりの自立語等を入れておく。

文法情報としては, 後続語及び先行語との関係を各々分類した後接情報及び前接情報(付表2)を情報コードとして与える。なお, 情報コードとして, 通常の品詞情報の外に, 文節頭あるいは, 漢字, 片仮名, 英字列といった字種等, 便宜的な分類を含んでいる。

3) 接続表 (付表3)

表は相互参照表の形式をとっており, 一方は前接情報(47種), 他方は後接情報(31種)から成っている。各情報コード間の接続の可否を示す外, 文節の切れ目を入れるべきか否かの情報を含める。

4.2 処理手順の概要

この処理手順を入力テキスト例(第1図)により説明する。その全体図を第2図に示す。

4.2.1 字種の判別

入力データの字種を先頭から1文字ずつ判定していき, 字種の変わり目で, その字種コード(下記の4種)とその位置(何字目から何字目まで)をスタックする。

字種は次の4種とし, 特殊記号は特殊記号処理ルーチンを通して4種の中のいづれかとみなす。

- ①句読点 ②平仮名 ③漢字, 片仮名, 英字 ④数字

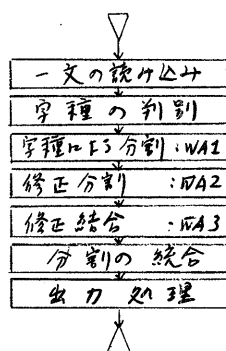
例文1 ③ 01, 10 ② 11, 11 ③ 12, 14 ② 15, 15 ③ 16, 19 ② 20, 22 ③ 23, 25 (

4.2.2 字種による分割処理

平仮名から非平仮名への変わり目(及び句読点の前後)をスタック WA1 に蓄積する。

例文1 従来粘土含有合成樹脂は粘土量が合成樹脂²⁰100重量部に対し20重量部以上になると、合成樹脂に対する粘土の分散性が悪くなり、又得られた粘土……

例文2 弗酸の1⁵重量%以下を珪弗酸で置き換えた弗酸-硝酸系酸洗液に0.01~5重量%のチオカルバゾン類及びチオカルバジド類の少なくとも1種を添加したことを特徴とする鋼材洗滌用組成物。



第1図 入力テキスト例

第2図 処理手順

例 例文1 11 15 26 28 39 45 48 51 55 57 65

例文2 03 11 15 17 20 30 40 50 58 64 67 74 79 87

4.2.3 修正分割処理

字種による分割では切れ残りが、正しく分割すべきものを、修正情報を用いて判定し、4.2.2と同じ形式でWA2に蓄積する。

例 例文1 02 36 61

例文2 48

例1の02, 61は漢字副詞表、例2は名詞接続表、例1の61は辞書検索による。

4.2.4 修正結合処理

字種による分割で分割されるが、結合するのが適当と判断されるものを判定し、WA3に蓄積される。

例 例文1 26 45 例文2 17

例文1はすべて辞書検索、例文2は複合動詞処理

4.2.3と4.2.4の処理は完全に独立した処理でなく、平行に行なわれる場合もある。又WA2とWA3の内容は必ずしもスタッフ・アドレスの上昇順になっていない。さうに同じものが混入することもあり得る。

4.2.5 分割の統合

WA3の分割点は、必ずWA2の中に含まれているはずであるからこれをWA1から消去する。修正されたWA1とWA2とを併合し、ソートを行ない、正しい分割箇所を得る。

4.2.6 出力処理

分割スタッフを参照して、入力文中に区切り符号を挿入し、グラフ・ディスプレイに出力する。

4.3 辞書検索手順

4.3.1 平仮名列の左端から最長優先で辞書検索を行ないPDS(I)に付くまで、同時に、その接続の可否をもスタッフする。残余の文字列に対し、これを右へくり返し行なう。

4.3.2 検索でなくなったところで、平仮名文字列の末尾に至ったかどうかを調べ、到達していれば、各PDS(I)の一番上の項をみて分割指示をWA2に蓄積する。

ハカ文字列	PDS (I)				
	1	2	3	4	5
変えられたことによって得られる	29 15 1				
減少されているため故障の	3 15 2	28 18 1	38 2 1	11 3 22 1	得 29 15 0
	20 15 1		11 3 22 1		
	3 15 2		11 3 22 1		
	4 6 1	7 8 2 0	11 3 22 1	ため 38 12 1	
	3 15 2	7 27 23 1	11 3 22 1	ため 3 15 2	
↑ 検索始点	↑ 文字列	↑ 結合値	↑ 結合値	↑ 結合値	
		結合値	1 (結合可)		
			2 (末尾となりうる)		
			0 (結合不可)		

第3図 辞書検索のPDSスタック例

4.3.3 末尾に至っていただければ、最後に検索した文字列が、文節の末尾となり得るものかどうかを調べ、可能な上と同様にする。不可ならば、 $Z = Z - 1$ とし、PDS (I) により引き返して処理を行なう。

4.3.4 辞書検索の段階で字種の異なる目を越えたかどうかを別にチェックし、越えていた結合情報をWA3に入れる。

上の例についての処理結果

変えることによって得られる

減少されているため故障の

↑ 文節区切り記号

4.4 複合動詞処理

修正結合処理のうち、複合動詞以外のものの処理は、平仮名処理ルーチン中で行なうが、複合動詞の処理は別個に行なう。

字種パターンが ① ② ③ ④ (口; 1字漢字 ○; 1字平仮名) の文字列を探索し、①が動詞連用形語尾 ②が動詞語尾であるならば、①と②の間を結合する。

例 丹酸と置き換えることがでる
電極面から洗い流されるが

メッキ浴から引き出し
心線の溶け込みが

5. 結果と考察

処理は、大型計算機 (TOSBAC 5600) 上で行ない。その操作は、その計算機に接続されているどのTSS 画像端末からでも行なうことができる。仕様概要を第2表に示す。処理の状況を逐次知ることができるよう、画面上に漢字仮名混じり文を単位に表示出力する。又モード指定をチェック、モードにあることにより、必要な部分のWA1, WA2, WA3の処理過程を表示出力する。一括処理は、入出力ファイルを指定することにより、高速処理を行なうことができる。

端末は、電話回線 (8,600 ボー) を介して接続されている。端末の画面上での漢字表示速度を上げるため、電総研で普通に使用されている漢字パターン (FONT 4000) を用いている。これは、32×32ドットをベースのストローク方式で集団ディスプレイ上に蓄積されている。

精度については、1回目の実験結果 (全テストの1/3) について分析を行なった。その結果92.5%が得られた。誤った部分について、現在表および辞書の修正、アルゴリズムの変更 (例; 平仮名末であっても、そこで切れてはいけない場合には引き返し処理を行なう) を行なっている。これにより精度は99%以上

第2表 システムの仕様

項 目	内 容
入出力媒体	オンライン・ユーザファイル(最大64万字)
処理単位	文(句点まで)
内部漢字コード	FONT K000 (JIS 第1水準) 10進4桁
プログラム言語	TSS FORTRAN
プログラムのサイズ	1200行, 占有領域 50KW (1W=36bit)
出力形式	TSS画像端末より操作し, 次のモードで出力 A) 入力文と分かつ書きた文の並び B) 分かつ書きた文の4出力 C) 分割過程 (WA1, WA2, WA3) を表示

前述の標本について, 実験を行なった結果

総文字数 274,650
 処理時間 1,340秒
 処理速度 205字/秒 (4.8ms/字)
 平均文字数 5.7字/文節

- 0498 此の\関係の\重要性は\、\高クロム合金に対し\約0.2%以下の\ケイ素含量を\保つことが\問題であるような\実質的な\ケイ素の\量を\、\一般に\本発明による\合金の\製造に\用いられる\重要原料が\含有していると言う事実を\基礎として\いる\、\
- 0499 若し\電解クロムの如き\低ケイ素原料を\用いるならば\、\合金の\価格は\甚だ\高価に\なる\、\
- 0500 第4図に\示した\資料の\知識から\、\良好な\耐蝕性\並に\経済的な\熔融作業と\両立する\クロムと\ケイ素の\含量を\選ぶことができる\、\
- 0501 \オーステナイトの中に\溶解し\、\溶体化焼鈍を\促進すると\されている\或成分を\添加することは\、\本発明の\範囲内に\あると\考えねばならない\、\
- 0502 その\理由は\、\硬度\、\強さ\、\及び\耐蝕性を\必要とするからである\、\
- 0503 Mo, Wと\Vの如き\成分は\此の\群に\属する\、\
- 0504 約732と\((113502)\)に\於ける\クリープ破断強さで\明らかなる\如く\、\此等の\成分の中で\、\Moが\最も\有望なものの\一つであると思われる\、\
- 0505 \全量で\約5%迄\この種の\成分を\添加することが\許される\、\
- 0506 \本発明の\合金は\、\酸化鉛の\存在下の\激烈な\酸化へ\抵抗性があるため\開発されてきたのであるが\、\他の\環境で\酸化に対する\抵抗性が\あるので\、\排気弁以外の\使用に\本合金は\適している\、\
- 0507 \本発明による\合金の\実施例を\第IX表に\示す\、\
- 0508 \表17\ \表21\ \上表に\記載した\合金は\いづれも\特許請求の\範囲に\開示する事項に\適合するものである\、\
- 0509 \本発明による\合金の\広い\範囲と\許容な\範囲は\次の如く\である\、\
- 0510 \表19\ \本発明の\実施態様は\次のとおりである\、\
- 0511 ((1)) \約20-70%ニッケル\、\4-20%マンガン\、\12-40%クロム\、\0.5%迄の\炭素\、\0.3%迄の\窒素\、\0.6%迄の\ケイ素\、\Mo, W\及び\Vの\群の中の\少なくとも\一つの\金属と\其等金属の\組合せの\全量が\5%迄\、\5%迄の\チタニウム\、\アルミニウムと\鋼の\各量が\4%迄\、\0.2%迄の\ホウ素\、\残余が\実質的な\鉄から\本質的に\成り\立っている\高耐蝕性合金\、\
- 0512 ((2)) \約0□-0.1%の\炭素\、\0.03-0.1%の\窒素\、\0.45%より\少ない\ケイ素\、\34-40%の\ニッケル\、\5-8%の\マンガン\、\20-36%の\クロム\、\Mo, W\及び\Vの\群の中の\少なくとも\一つの\金属と\其等金属の\組合せが\0\または\0□-3%\、\1.5-3%の\チタニウム\、\0.2-1.5%の\アルミニウム\、\0\または\0□-0.5%の\鋼\、\0\または\0□-0.1%の\ホウ素を\含有す\ \る\第((11))\項の時\焼硬化性合金\、\

第4図 分かつ書まれた出力例

が得られるとおもわれる。しかし、なお残された1%の内容は、表や辞書に登録し、その処理を行なうために、処理前は正しかったものを誤りとする場合が大部分である。例えば、前、時、中といった文字を形式名詞として処理した場合、送り及びといった文字連系を複合動詞の処理を行なった場合などがある。このような場合、現在の方法では、出現率等によりその処理方法を決定せざるを得ない。

詳細な分析結果については、別の場合に述べるが、精度の評価には、文節をどのように定義するかにある。また、実際にこの結果を用いてテキスト分析をやることによって、このシステムの評価は決まる。

システムの汎用性という面から、今後他の標本、別の分野の標本に対する実験を行なって、修正情報について検討を加える予定である。

最後に、本研究の機会を与えて下さった石井治ソフトウェア部長、島居栄次言語処理研究室長、又実体調査を行ない、プログラムを作成して下さい、IBSの木村睦子さん、佐々木幸博さんに感謝いたします。

参考文献

- 1) 夢沼良一、山崎敏子；分かつ書き一案、機械翻訳研究会、1965.5
- 2) 江川清；漢字から混り文の「自動単位分割」に関する研究、計量国語学、43/44号、1968
- 3) 江川清；単位分割自動化のシステムについて、計量国語学、51号。
- 4) 石綿敏雄他；単語認定プログラム、CL研究委員会、情報処理学会、1969.3.
- 5) 小野寺夏生他；化学的に有用なキーワードの検索のための自動助詞分割システム、第14回情報科学技術研究集会論文集、JICST.
- 6) 秋元啓次他；日本特許情報商標に於ける分かつ書きの二、三の実施例、第6回情報科学技術研究集会発表論文集、JICST.
- 7) 田中康仁、古賀勝夫；日本文の自動分かつ書きについて、第15回情報科学技術研究集会発表予稿、JICST. 1978.10
- 8) 坂本義門；文節単位の自動分割法—字種と平仮名連系による—、計量国語学会、
- 9) 坂本義門；ALPS-CONCORDANCEの生成と検索システムについて、第14回情報科学技術研究集会発表論文集、JICST.
- 10) 婦人雑誌の用語 P18~27、国立国語研究所報告46、1972

付表 1-A 漢字副詞表

一度	且	充分	各々	従来品	普通鋼	短時間	適宜
一定時間	今一度	再度	多少	従来形式	普通	種々	鋭意
一層	以下	別途	多数条	従来技術	極力	結局	長年
一方側	以上	十分	多数	従来方法	毎秒	至極	長時間
一方	他方	又	夫々	従来法	比較的	近年	順次
一昼夜	何等	最初	少量	従来	漸次	近時	尚*
一旦*	依然	若干	幾分	微量	現在	通常	勿論

*印は非当用漢字

付表 1-B 名詞接続詞表

且つ	例えげ	即ち	又付	及び	並に	並びに	或いは	或は
----	-----	----	----	----	----	-----	-----	----

付表 1-C 助数詞表

%以下	%迄	倍強	倍	回転	時間	週間		
%以上	以下	倍程度	分以上	回	秒以上	重量部		
%程度	以上	倍量程度	分間	時間以上	秒間			

種 別	文 字
句切り符号	、, ., …, 空白, @
かつ二類	(,), [,], 「, 」, “, ”,
演算子・結合子	+, ×, ±, <, >, ≤, ≥, →, ~, /, =
助数詞, 添字	°, ', ", °C, %, A, ↓, ギリシア文字, 添数字, 肩つき+
ギリシア字,	肩つき-, 肩つき 〇~?, 卩, ローマ数字, (1)~(9),
ローマ数字	cm, mm, kg, mg, m/m, l, ml, m ² , m ³ , cm ³ ,
その他	No., °F, No. ④, (株)
多用途記号	-, . (ピリオド), , , ° (中黒), ;,

付表 2-A 前接情報分類表

コード	内 容	コード	内 容
1	始発状態	10	の (格助詞)
2	名詞 (複合語の要素にナリうるもの)	11	係助詞 け, も, ため, こそ, しか
3	上記以外の自立語	12	副助詞 (なご, はかり, たけ…), と共に, に従って
4	送りがおねと, 漢字の直後のひたくるもの	13	まで (格助, 副助)
5	か, かな, かどうか, うい, だろう, であろう, なじげ	14	から (格助)
6	だ, だった, で, である, ではない, ではない	15	が (格助), を
7	に	16	に連れて
8	と, へ, や, として, において, について, によつて…	17	より
9	という, といった	18	な (形動語尾)

19	すぎる, そう (推定)	34	と (接助, 引用)
20	カ変語尾	35	接続助詞 として, から, か, けれど, こ, どの
21	于世	36	そう (伝聞)
22	形容詞語尾	37	とす
23	せ, れ	38	ごと
24	ち (打消), ん (む)	39	形式名詞
25	た	40	まい
26	た	41	よう
27	て	42	べし, べく, に当り, 毎, た (禁止)
28	ろ, ろ, れは, よ	43	の (媒体), ので, のに, にもかかわらず
29	つつ, ながら, たい, 動詞性, 形容詞性接辞	44	う, り
30	み (一段語尾, 補助動詞)	45	なり (補助動詞)
31	す (名詞性接辞)	46	補助動詞 しての後に来るもの
32	しめ	47	のみ, だけ, ばかり (助詞に後続するもの)
33	られ		

付表 2-B 後接情報分類表

コード	内 容	コード	内 容
1	始発状態	17	五段動詞連用形 (サ行), カ変連用形
2	名詞	18	動詞, 形容詞終止連体形
3	形容動詞語幹	19	命令形
4	カ変動詞語幹	20	カ変連用形 こ, じ
5	形容詞語幹	21	助動詞 た
6	五段動詞未然形	22	形容詞語尾 く
7	五段連用イ音便 (清音), 促音便	23	助動詞 て
8	カ変未然形語尾 せ, せ	24	べき
9	た (断定)	25	接続助詞 から
10	う, よう, まい	26	格助詞 と
11	文節末にしかたなり得ないもの	27	へ, から, より, で, において
12	副詞的名詞	28	まで
13	副詞	29	に
14	連体詞, 格助詞 (の)	30	15 + 18 (「い」のため)
15	一段活用未然連用語尾	31	18 + 22 (「く」のため)
16	五段動詞連用形 ①		

付表 3 接 続 表

[illegible]