

国立国語研究所学術情報リポジトリ

An automatic processing of conjugation in old Japanese and modern Japanese

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 麿岡, 昭夫, TURUOKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001022

文語形・口語形活用語の 代表形への変換処理について

霧 岡 昭 夫

①・はじめに

現在、われわれは「漱石・鷗外の用語の研究」のために、用語検索システムを開発中である。このシステムでは、電子計算機を使用するので、プログラム・付加情報コード・入出力の機器等、前の「新聞の語彙調査」で用いたものは利用できる限り生かすという方針をとっている。

しかし、漱石・鷗外の作品の用語と新聞のそれとでは、表記・語形・語法などで質的な相違がある。したがって、語彙調査で開発した処理方法を応用することができない場合が出てくる。

ここで活用語の代表形への変換処理についてとりあげるのは、前に同様な目的をもって「新聞の語彙調査」で開発された「代表形変換ルーチン」^(注1)が、用語検索（ここでは特に索引作り）には適していないからである。

明治期の文献は、一応歴史的かなづかいで書かれてはいるが、そのかなづかいは必ずしも統一されておらず、乱れているもの、後世書き直されたものが相当ある。また、活用語の語形も、口語形・文語形のいずれも多く用いられている。

原文に現われた活用語が、新かな・旧かなのいずれであっても、平かなでも片かなでも、漢字だけで「居(た)」「来(ます)」のように表記されていても、また、口語形であろうと文語形であろうと、すべてその代表形に変換されるような処理方式を開発しようとするのが本研究の狙いである。

(注1) 江川清『「活用形処理」の自動化における一方式』(国立国語研究報告34「電子計算機による国語研究・II」)

1・対象

このプログラムは、「漱石・鷗外の用語の研究」用に開発したものであるが、単語として単位が切れていて、フォーマット・付加情報がこのシステムと同じであればどんな作品の用語であっても利用できる。このシステムで用いられているデータのフォーマットはつぎのとおりである（カッコ内の数字は桁数）。

(I 図)

単位	見出し	読みがな	情報	ページ	行	文種	題	段落	エラー	E/i
(3)	(20)	(20)	(6)	(4)	(2)	(1)	(1)	(4)	(2)	(1)

入力データ
(固定長64桁)

ここで行なう処理は、上の入力データ末尾 (E/i マークの前) に、代表形見出し (20ケタ)^(注2) と代表形読みがな (20ケタ) との欄を付け加え、それぞれの欄に見出し、読みがなを代表形に変換したものを入れるようにするものである。代表形見出し、代表形読みがなは、活用語においては終止形とし、活用語以外では見出し、読みがなと同じものとする。

ここでは、情報 (語種・品詞・活用形・活用行・連語・スペースの6桁) が付加されたデータを用いる。この研究は、そのうちの、品詞・活用形・活用行の各情報を利用して活用語を代表形に変換しようとするものである。

活用語は、動詞・動詞性接辞・形容詞・形容詞性接辞・助動詞 (形容動詞語尾を含める) の五種類であるが、この研究では、その五つのうち、助動詞の処理はとりあげない。それは、助動詞の場合、その詞数が多く、活用型式もさまざまであるため、処理ステップが長くなるうえ、活用テーブルも大きなものを必要とするので、現在の電子計算機のメモリー量ではこなせないからである。そこで、今までの所では助動詞の処理はせず、動詞・動詞性接辞・形容詞・形容詞性接辞の処理だけ行なうようなプログラムにした。ただし、助動詞の処理専用のプログラム (別にランを行なう) を作ることを考えているし、またメモリーの調整ができれば今のプログラムにあとからでも組みこめるようにしてあるので、近い将来は助動詞の処理を行なうことが可能になるはずである。

また、活用語の中でも、「なすって」「見りゃ」等、活用の形が明確でない

ようなもの、他と融合しているものなどは、代表形が決められないので、これも処理を行なわない。

したがって、今代表形変換処理の行なわれるデータは、動詞・動詞性接辞・形容詞・形容詞性接辞の、規則的な活用型（四段・上一段・下一段・上二段・下二段・変格等）を有し、代表形が一義的に定まるものである。

（注2）見出し、読みがな、代表形見出し、代表形よみがなは、20ケタで、データがそれより短いときは、後ろの残った部分にはスペース(Ⓢ)がはいる。なお、計算機コード20の桁は、漢字テレタイプコードで10文字分である。

2・代表形を一義的に定める方法

本稿0で述べたように、この代表形処理は、かなづかい・活用型式について新旧かなづかい、口語文語活用型式のいずれをもおあったものにするのが目的である。すなわち

- ①旧かな表記・文語形（「恋ふる」の類）
- ②旧かな表記・口語形（「植ゑる」の類）
- ③新かな表記・文語形（「恋うる」の類）
- ④新かな表記・口語形（「植える」の類）

の四通りの組み合わせを一つの処理ですませようということである。すると、「変へ」は①に属するか②に属するか、「変え」は③に属するか④に属するか「行き」はどこに属するのか、といった問題が出て来る。したがって、上の四類のものを一義的に処理するアルゴリズム（演算規則）を立てておく必要がある。

そこで、とりあえず、下の表1に示すような用言の変化形の整理を試みた。この表の1から27までの変化形は、つぎの手続きによって抽出されてきたものである。

<一般形>

- | | |
|--------|-------|
| 1・～ナイ | 5・～ウ |
| 2・～ズ | 6・～ヨウ |
| 3・～セル | 7・中止法 |
| 4・～サセル | 8・～タ |

9・(動詞については) ～マス	12・～バ
(形容詞については) ～ゴザイマス	13・裸命令
10・言い切り	14・～ロ
11・連体用法	15・～ヨ

<特殊形>

21・～シム	(1～6に同形のあるときは無記入)
22・～ケリ	(7～9 “ ”)
23・言い切り	(10・11 “ ”)
24・～メリ	(10・11 “ ”)
25・連体用法	(10・11 “ ”)
26・～ド	(12 “ ”)
27・命令形	(13～15 “ ”)

なお、表の中の各種記号は、つぎのとおりである(詳細は後述)。

×……変化形(活用形)なし

／……活用情報統合

ハダカ英文字……活用情報移行

○囲みの英文字……活用種別移行

⊕……語幹接続

→……参考項目(相補的關係)

以下、この表の内容について、若干の説明を加えることにする

ア) 一般形と特殊形

この表の一般形と特殊形は、主として口語活用と文語活用の対応であるが、①②の一般形(ワア五)と特殊形(ハ四)には新旧かなづかいの対応がからんでいない。

特殊形の21～27の「変化形」の欄のブランクは、一般形と共通の変化形が抽象されるもので、これについては、一般形において処理することにした。

なお、特殊形の情報欄に→印のあるものは、単に参照項目を示したものであり、一般形に見当たらない変化形は→印の個所に出ていることを表わしている。たとえば②の「→ザ変B・感ず・信ず」の場合、一般形にあげた「Gざ・ザ上

一・閉じる・感じる」に示されていない変化形は、「ザ変B」を見よという意味である。

イ) 情報

「情報」欄に表示してある「Fワ」「Kか」「M」などの活用情報は、必ずしも、国文法でいう「活用の種類」に対応していない。たとえば、国文法で「タ上二」とされる「落ちケリ」の活用情報も、「落ちマス」の「落ち(タ上一)」の活用情報も、ともに「Gた」の形になる。普通「ヤ下二」とされる「越えズ」の「越え」も、「ア下一」とされる「越えヨウ」の「越え」も、活用情報は、ともに「Iあ」に統一されている。これは、一つの語形について、代表形(表の10, または23にあたる)が、一義的に決まるように配慮してあるからである。

ウ) 種別と例語

「種別」欄に示した「ワア五」とか、「カ変」とか、あるいは「ク活」というような表示は、「例語」として掲げた用言が、国文法でどう扱われているかを示したものである。これは、表を見やすくするためだけのもので、「活用情報」や、変化形として記入されている語尾とは、必ずしも対応しない——例えば「ヤ上一」と表示してある㊸の「報いる・老いる」が、活用情報は、「Gあ」となっていることなどはその一例である。また、「ザ変」「シク活」といったような種別も設けてある。

「種別」欄のA・B・C・Dの記号は、Aには、初歩的な文法で扱われるような代表的な語の活用を示し、B以下には、特別な活用のある語あるいは文語扱いにされているものなどを掲げてある。

エ) 変化形の示し方

この表の変化形(1~27まで)の抽出のしかたは前に述べてあるが、これによって抽出される語形変化が、「種別」欄のA B C Dで示すように、2系列以上にわたるときは、同一形は示さないで、㊸のような符号で掲げた。㊸の「Fワ(ワア五A)言う・舞う」と㊹の「Eワ(ワア五B)問う・乞う」を例にとると、「問う・乞う」の系列、すなわちBの系列は、第8変化形「ウ」が、A系列と異なるのみで、その他は系列と同形であるので㊸が表示してある。つき

表 1

番 号	一 般 形													
	情報	種 別	例 語	1	2	3	4	5	6	7	8	9	10	11
①	Fワ	ワA	言う・舞う	ワ	ワ	ワ	×	オ	×	イ	ッ	イ	ウ	ウ
②	Fワ	B	問う・乞う	Ⓐ	Ⓐ	Ⓐ	×	Ⓐ	×	Ⓐ	ウ	Ⓐ	Ⓐ	Ⓐ
③	Fか	カ五A	咲く・開く	カ	カ	カ	×	コ	×	キ	イ	キ	ク	ク
④	Fか	B	い(行)く	Ⓐ	Ⓐ	Ⓐ	×	Ⓐ	×	Ⓐ	ッ	Ⓐ	Ⓐ	Ⓐ
⑤	Fか	C	ゆ(行)く	Ⓐ	Ⓐ	Ⓐ	×	Ⓐ	×	Ⓐ	×	Ⓐ	Ⓐ	Ⓐ
⑥	Fが	ガ五	注ぐ・脱ぐ	ガ	ガ	ガ	×	ゴ	×	ギ	イ	ギ	グ	グ
⑦	Fさ	サ五A	話す・越す	サ	サ	サ	×	ソ	×	シ	シ	シ	ス	ス
⑧	Fさ	B	訳す・愛す	Ⓐ	Ⓐ	Ⓐ	K	Ⓐ	K	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
⑨	Fさ	C	任す	Ⓐ	Ⓐ	Ⓐ	I	Ⓐ	I	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
⑩	Fさ	D	待たす・合わす	Ⓐ	Ⓐ	Ⓐ	×	Ⓐ	×	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
⑪	Fた	タ五	立つ・待つ	タ	タ	タ	×	ト	×	チ	ッ	チ	ツ	ツ
⑫	Fな	ナ五	死ぬ	ナ	ナ	ナ	×	ノ	×	ニ	ン	ニ	ヌ	ヌ
⑬	Fば	バ五	呼ぶ・飛ぶ	バ	バ	バ	×	ボ	×	ビ	ン	ビ	ブ	ブ
⑭	Fま	マ五	進む・富む	マ	マ	マ	×	モ	×	ミ	ン	ミ	ム	ム
⑮	Fら	ラ五A	乗る・取る	ラ	ラ	ラ	×	ロ	×	リ	ッ	リ	ル	ル
⑯	Fら	B	蹴る	Ⓐ	Ⓐ	Ⓐ	I	Ⓐ	×	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
⑰	Fら	C	なさる・下さる	Ⓐ	Ⓐ	×	×	Ⓐ	×	Ⓐ	Ⓐ	イ	Ⓐ	Ⓐ
⑱	Fら	D	ある	×	Ⓐ	Ⓐ	×	Ⓐ	×	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
⑲	Gあ	ア上一	射る・報いる	イ	イ	×	イ	×	イ	イ	イ	イ	イル	イル
⑳	Gか	カ上一	着る・飽きる	イ	キ	×	イ	×	イ	キ	キ	キ	キル	キル
㉑	Gが	ガ上一	過ぎる・よぎる	ギ	ギ	×	ギ	×	ギ	ギ	ギ	ギ	ギル	ギル
㉒	Gぎ	ザ上一	閉じる・感じる	ジ	ジ	×	ジ	×	ジ	ジ	ジ	ジ	ジル	ジル
㉓	Gた	タ上一	落ちる・満ちる	チ	チ	×	チ	×	チ	チ	チ	チ	チル	チル
㉔	Gだ	ダ上一	恥ぢる・閉ぢ	ヂ	ヂ	×	ヂ	×	ヂ	ヂ	ヂ	ヂ	ヂル	ヂル
㉕	Gな	ナ上一	似る・煮る	ニ	ニ	×	ニ	×	ニ	ニ	ニ	ニ	ニル	ニル
㉖	Gは	ハ上一	干る・強ひる	ヒ	ヒ	×	ヒ	×	ヒ	ヒ	ヒ	ヒ	ヒル	ヒル
㉗	Gば	バ上一	伸びる・わびる	ビ	ビ	×	ビ	×	ビ	ビ	ビ	ビ	ビル	ビル
㉘	Gま	マ上一	見る・染みる	ミ	ミ	×	ミ	×	ミ	ミ	ミ	ミ	ミル	ミル
㉙	Gあ	ヤ上一	報いる・老いる	/	/	×	/	×	/	/	/	/	/	/
㉚	Gら	ラ上一	下りる・こりる	リ	リ	×	リ	×	リ	リ	リ	リ	リル	リル
㉛	Gわ	ワ上一	居る・率る	キ	キ	×	キ	×	キ	キ	キ	キ	キル	キル
㉜	Iあ	ア下一	得える・越える	エ	エ	×	エ	×	エ	エ	エ	エ	エル	エル
㉝	Iか	カ下一A	向ける・受ける	エ	ケ	×	エ	×	エ	ケ	ケ	ケ	ケル	ケル
㉞	Iか	B	蹴る	F	Ⓐ	F	Ⓐ	F	×	F	F	F	F	F
㉟	Iが	ガ下一	曲げる・逐げる	ゲ	ゲ	×	ゲ	×	ゲ	ゲ	ゲ	ゲ	ゲル	ゲル
㊱	Iさ	サ下一A	失せる	セ	セ	×	セ	×	セ	セ	セ	セ	セル	セル
㊲	Iさ	B	任せる	Ⓐ	Ⓐ	F	Ⓐ	F	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ

					特 殊 形								
12	13	14	15	情報	種 別	例 語	21	22	23	24	25	26	27
エ Ⓐ	エ Ⓐ	×	エ Ⓐ	Fは	ハ 四	言ふ・乞ふ	ハ	ヒ	フ	フ	フ	へ	へ
ケ Ⓐ	ケ Ⓐ	×	ケ Ⓐ										
ゲ Ⓐ	ゲ Ⓐ	×	ゲ Ⓐ										
セ Ⓐ	セ Ⓐ	×	セ Ⓐ	→	サ変D	愛す・訳す							
テ Ⓐ	テ Ⓐ	K	I Ⓐ										
ネ Ⓐ	ネ Ⓐ	×	I Ⓐ										
ベ Ⓐ	ベ Ⓐ	×	I Ⓐ		Kな	ナ 変					ヌル	ヌレ	
メ Ⓐ	メ Ⓐ	×	I Ⓐ										
レ Ⓐ	レ Ⓐ	×	I Ⓐ	→	カ下	蹴る							
ラ Ⓐ	ラ Ⓐ	×	Ⓐ	Fら	ラ 四	なさる・下さる							レ
ラ Ⓐ	ラ Ⓐ	×	Ⓐ	Kら	ラ 変	有り・居り				リ			
イレ	×	イ	イ	Hあ	ア上二	用う・老う			ウ	ウ	ウル	ウレ	
キレ	×	キ	キ	Hか	カ上二	尽く・飽く			ク	ク	グル	クレ	
ギレ	×	ギ	ギ	Hが	ガ上二	過ぐ・風ぐ			グ	グ	グル	グレ	
ジレ	×	ジ	ジ	→	ザ変B	感ず・信ず							
チレ	×	チ	チ	Hた	タ上二	落つ・満つ			ツ	ツ	ツル	ツレ	
ヂレ	×	ヂ	ヂ	Hだ	ダ上二	恥づ・閉づ			ヅ	ヅ	ヅル	ヅレ	
ニレ	×	ニ	ニ										
ヒレ	×	ヒ	ヒ	Hひ	ハ上二	恋ふ・強ふ			フ	フ	フル	フレ	
ビレ	×	ビ	ビ	Hび	バ上二	伸ぶ・化ぶ			ブ	ブ	ブル	ブレ	
ミレ	×	ミ	ミ	Hま	マ上二	浴む・染む			ム	ム	ムル	ムレ	
/	×	/	/	Hや	ヤ上二	報ゆ・悔ゆ			ユ	ユ	ユル	ユレ	
リレ	×	リ	リ	Hら	ラ上二	降(お)る			ル	ル	ルル	ルレ	
キレ	×	キ	キ										
エレ	×	エ	エ	Jあ	ア下二	得(う)・加う			ウ	ウ	ウル	ウレ	
ケレ	×	ケ	ケ	Iか	カ下二	受く			ク	ク	クル	クレ	
F	F	Ⓐ	Ⓐ	→	カ下	蹴る							
ゲレ	×	ゲ	ゲ	Jが	ガ下二	告ぐ・曲ぐ			グ	グ	グル	グレ	
セレ	×	セ	セ	Jさ	サ下二	乗す・寄す			グ	グ	スル	スレ	
Ⓐ	F	Ⓐ	Ⓐ	→	サ 五	任す			ス	ス			

番号	一 形 形														
	情報	種 別	例 語	1	2	3	4	5	6	7	8	9	10	11	
38	I	ぎ	ザ下	混ぜる・はぜる	ゼ	ゼ	×	ゼ	×	ゼ	ゼ	ゼ	ゼ	ゼ	ゼ
39	I	た	タ下	建てる・当てる	テ	テ	×	テ	×	テ	テ	テ	テ	テ	テ
40	I	だ	ダ下	出る・なでる	デ	デ	×	デ	×	デ	デ	デ	デ	デ	デ
41	I	な	ナ下	寝る・まねる	ネ	ネ	×	ネ	×	ネ	ネ	ネ	ネ	ネ	ネ
42	I	は	ハ下	考へる・経へる	へ	へ	×	へ	×	へ	へ	へ	へ	へ	へ
43	I	ば	バ下	並べる・延べる	べ	べ	×	べ	×	べ	べ	べ	べ	べ	べ
44	I	ま	マ下	求める・醒める	メ	メ	×	メ	×	メ	メ	メ	メ	メ	メ
45	I	あ	ヤ下	越える・燃える	/	/	×	/	×	/	/	/	/	/	/
46	I	ら	ラ下	恐れる・倒れる	レ	レ	×	レ	×	レ	レ	レ	レ	レ	レ
47	I	わ	ワ下	植ゑる・据ゑる	エ	エ	×	エ	×	エ	エ	エ	エ	エ	エ
48	K	か	カ 変	来る	コ	コ	×	コ	×	コ	キ	キ	キ	クル	スル
49	K	さ	サ変A	する	シ	セ	サ	セ	×	シ	シ	シ	シ	スル	ズル
50	K	さ	B	す	㊤	㊤	㊤	㊤	×	㊤	㊤	㊤	㊤	ス	㊤
51	K	さ	C	決する・愛する	㊤	㊤	F	㊤	F	㊤	F	F	F	㊤	㊤
52	K	さ	D	決す・私す	㊤	㊤	F	㊤	F	㊤	F	F	F	F	㊤
53	K	ぎ	ガ変A	命ずる・信ずる	G	ゼ	×	ゼ	×	G	G	G	G	ズル	ズル
54	K	ぎ	B	命ず・信ず	G	㊤	×	㊤	×	G	G	G	G	ズ	㊤
55	M		形容A	赤い・寒い	ク	カラ	×	×	カロ	×	ク	カッ	ウ	イ	イ
56	M		B	遠い・多い	㊤	㊤	×	×	㊤	×	㊤	㊤	㊤	㊤	㊤
57	M		C	正しい・美しい	㊤	㊤	×	×	㊤	×	㊤	㊤	㊤	㊤	㊤
58	M		D	同じ	×	ジカ ラ	×	×	ジカ ロ	×	ジク	×	×	ジ	ジ
59	V	ん	変則活用												
60	Z	ん	活用不明												

に、この表では、「語幹語尾」の形が同一になるものは、たての系には重複しないようにしてある。たての系（1～27の各行）の一行に、同形の「語幹＋語尾」が生じうる場合には、いずれか一つのみを生かしてその活用情報のほか英文字を、他方の欄に記入する。たとえば、⑨の「Fさ（サ五C）」の第4形「任かせサセル」「任かせラレル」などの「任かせ」の形であるが、これは⑦の「Iさ（サ下一B）」の第4形「任かせ」と重複する。そのため⑨の「Fさ（サ五）」の第4形にはIを記入し、その処理が、「Fさ」から「Iさ」に移されることが示してある。したがって、「任かす」の第4変化形「任せ」の活用

					特 殊 形								
12	13	14	15	情報	種 別	例 語	21	22	23	24	25	26	27
ゼレ	×	ゼ	ゼ	Jざ	ザ下二	混(ま)ず			ズ	ズ	ズレ	ズレ	
テレ	×	テ	テ	Jた	タ下二	当つ・捨つ			ツ	ツ	ツレ	ツレ	
デレ	×	デ	デ	Jだ	ダ下二	出づ・賞づ			ヅ	ヅ	ヅレ	ヅレ	
ネレ	×	ネ	ネ	Jね	ナ下二	重ぬ・寝ぬ			ヌ	ヌ	ヌレ	ヌレ	
ヘレ	×	ヘ	ヘ	Jは	ハ下二	考ふ・経(ふ)			フ	フ	フレ	フレ	
ベレ	×	ベ	ベ	Jば	バ下二	延ぶ・述ぶ			ブ	ブ	ブレ	ブレ	
メレ	×	メ	メ	Jま	マ下二	求む・嵌む			ム	ム	ムレ	ムレ	
/	×	/	/	Jや	ヤ下二	越ゆ・覚ゆ			ユ	ユ	ユレ	ユレ	
レレ	×	レ	レ	Jら	ラ下二	恐る・晴る			ル	ル	ルレ	ルレ	
エレ	×	エ	エ	Jあ	ワ下二	植う・裾う							
クレ	コイ	×	コ	Kか	カ 変	来(く)			ク	ク			
スル	セイ	シ	セ										
Ⓐ	Ⓐ	Ⓐ	Ⓐ										
Ⓐ	Ⓐ	Ⓐ	F										
Ⓐ	Ⓐ	Ⓐ	F	→	サ 五	訳す・愛す							
ズレ	×	G	ゼ										
Ⓐ	×	G	Ⓐ										
ケレ	カレ	×	カレ	N	ク 活	寒し・遠し		カリ	シ	カル	キ	カレ	
Ⓐ	Ⓐ	×	Ⓐ									シカレ	
Ⓐ	Ⓐ	×	Ⓐ	N	シク活	正し・美し		シカリ	シ	シカル	シキ	シカレ	
×	N	×	×	N	シク活	同じ		シカリ	(シ)	シカル	シキ	シカレ	

情報は「Iさ」となる。

⑤の「M(形容B) 遠い・多い」の第9変化形欄に記入してある⊕印は、「とお・ゴザイマス」「おお・ゴザイマス」のように、活用語尾のない変化形を示したものである。

また、さきに(イ)(ウ)の各項で説明した、「Gあ(ヤ上一) 報いる・老いる」「Iあ(ヤ下二) 越える・燃える」の変化形欄には/ (斜線) が引いてある。これは、無条件に、「Gあ」「Iあ」として処理することを示すものである。

エ) 変則活用と活用不明

「変則活用」と「活用不明」には、それぞれ活用情報「Vん」「Zん」が与えてある。変則活用というのは、「すいマセン」の「すい」, 「歩っテイル」の「歩っ」, 「それ見い」の「見い」など、この表に収容していない変化形である。方言独特の活用形なども含まれる。しかし、「Vん」が与えられるのは、すべて用言（動詞・形容詞）の範囲に属するものについてのみである。もし、用言以外のものがまぎれこんだり、全く意味不明のデータが出て来た場合には、「Zん（活用不明）」として処理される。



上の表は、先にも述べたように、用言について、その代表形を一義的に定めるために作ったものである。すなわち、用言の変化形はすべて、上の表にもとづきただ一つの活用情報が与えられ、一般形であれば10の形が、特殊形であれば23の形が代表形とされるのである。このために上の表は、用言の、口語・文語および新旧かなづかひのすべてについて、生じうる変化形を網羅することを第一の目的としている。

上の表はしたがって、新かなづかひの文章でも旧かなづかひの文章でもまたその混乱している文章でも処理することを可能にし、口語形だけの文——実際には多少文語形がまざる場合がふつうであるが——や、口語形・文語形混合文の処理ができる。文語形だけの文、すなわち古文の処理はこのままでは、口語形の代表形が出てしまうので適しないが、このあと口語形の代表形を文語形に直す処理システムを開発しておけば、いわゆる文語文法にのった、代表形変換処理を行なうことも可能になる。

3・代表形変換の方法

本システムのデータには、品詞や活用などの情報が付加されている。すなわち、1のI図で「情報」となっている六ケタの部分である。

その情報のうち、2番目（品詞情報）、3番目（活用情報）、4番目（活用情報）の三つの情報を利用し、活用語尾テーブルを用いて代表形変換を行なう方式をここでは取る。

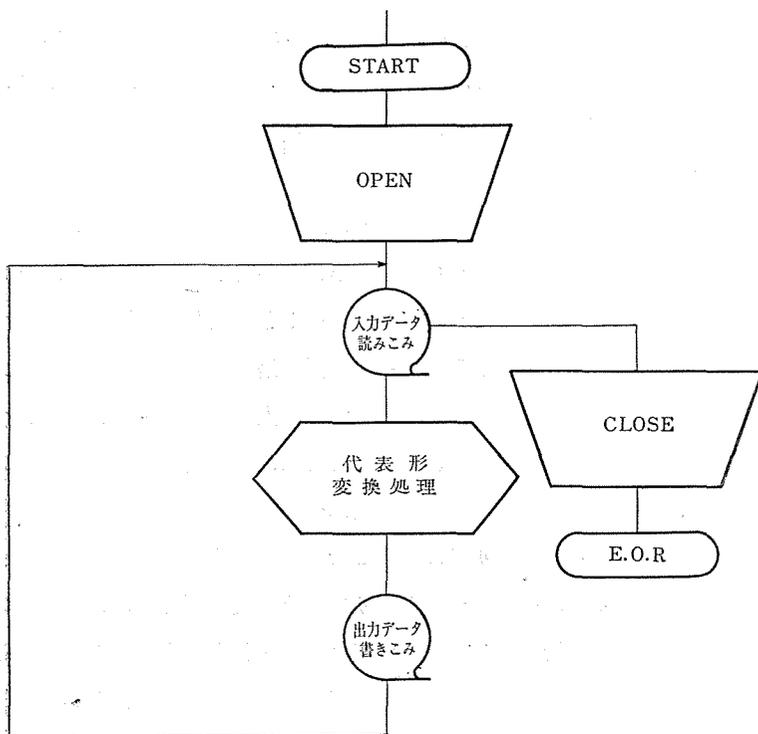
品詞コードは品詞により付加されたもので、ここではE（動詞）+（動詞性

接辞) M (形容詞) N (特殊活用形容詞) - (形容詞性接辞) のいずれかを有するデータを語尾テーブルを用い、後述の方法をもって代表形に変換する。その変換処理のアウトラインはII図の、出力データはIII図のようになる。なお、その処理方法は、似たものをまとめて次の五とおりにしてある。

(品詞情報・活用形情報) (語尾の字数)

- | | | |
|-----------------|--|-----|
| (1) 五段・四段活用処理 | EF・+F | 1 |
| (2) 上一段・下一段活用処理 | EH・EI・+G・+I | 1~2 |
| (3) 上二段・下二段活用処理 | EH・EJ・+H・+J | 1~2 |
| (4) 変格活用処理 | EK・+K | 1~2 |
| (5) 形容詞活用処理 | M [Ⓢ] ・N [Ⓢ] ・- [Ⓢ] | 0~3 |

上で、形容詞には活用形情報を付けていないのでその欄には[Ⓢ](スペース)がはいっている。



(II図)

(Ⅲ図) 出力データ (固定長104桁)

単位	見出し	よみがな	情報	ページ	行	文種	題	段落	エラー	代表形見出し	代表形よみがな	E/i
(3)	(20)	(20)	(6)	(4)	(2)	(1)	(1)	(4)	(2)	(20)	(20)	(1)

() 内の数字は桁数

また、動詞および動詞性接辞の情報欄では、活用形情報のあとに活用行情報がはいっている。これは、表の「情報」欄の、後半のかな文字にあたる。形容詞、形容詞性接辞については、この欄にもスペースがはいっている。

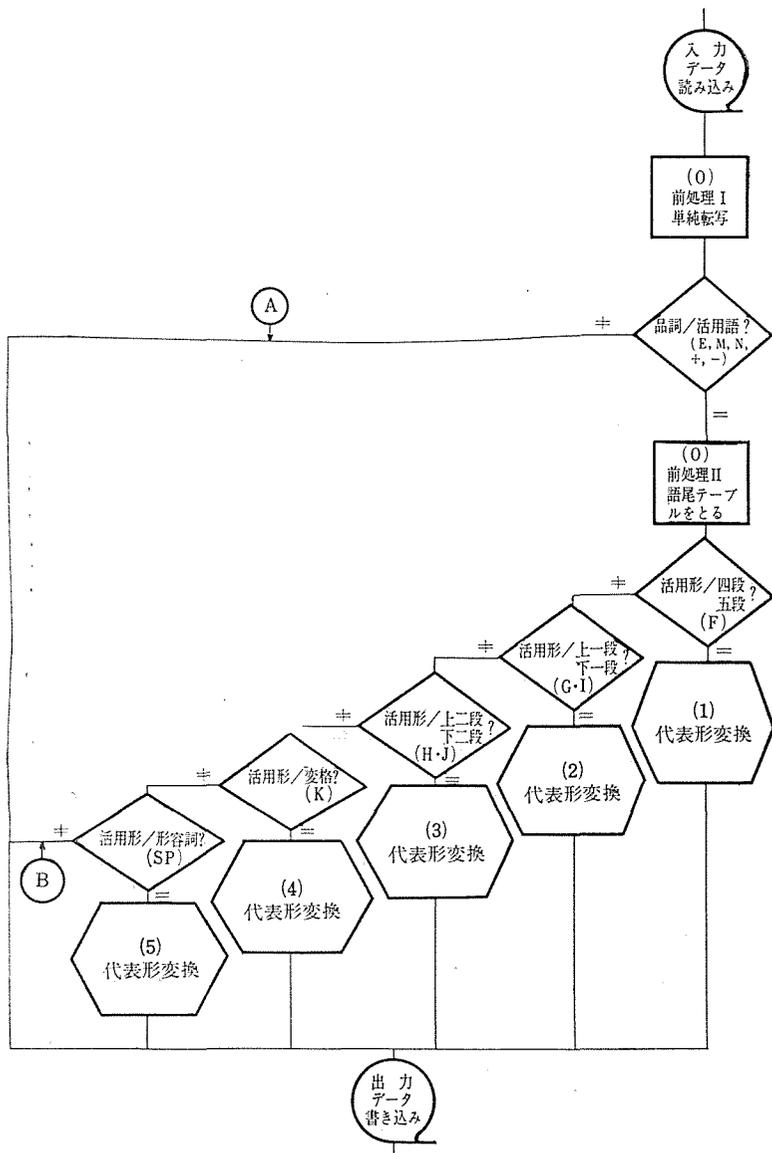
代表形変換のシステムはⅣ図に示したようになる。以下に、代表形変換の理論を模式的に述べる。現在用いられている電子計算機のコードには漢字・かなが無いので、実際のデータの、見出し、読みがな、代表形見出し、代表形読みがなの欄では、漢字テレタイプコード（電子計算機コード二列で一字を示す）を用い、活用行情報ではワア行→0、あ行→1、か行→2のように数字に変換したデータを用いている。しかし、変換の原理を説明するには、機械のコードで述べるよりもかなはかなで、漢字は漢字で書いておく方が理解しやすいと考えられる。したがってここで模式的に論ずるのである。

(O) 前処理

本システムで用いられるデータは、一応、品詞別活用別などのファイルにソートされてはいないもの、すなわち原文の単語の順番どおりに配列されたデータを用いるようになっている（もちろん、ソートされたものの処理も行なえる）。したがって、読みこんだ入力データが活用語であるか、非活用語であるかでその処理が異なる。また活用語であっても、上の(1)～(5)のいずれに属するかでまた、処理方法が変わる。ここでは、前処理としてそれぞれの過程において、すべての場合に共通する処理について述べる。

まず入力データをリードエリアに読みこむと、「単位」から「エラー」までの63桁をライトエリアに移す。ライトエリアのその後の部分へ、入力情報の見出し、読みがなの各20桁を移し（仮の代表見出し・代表読みがなとする）そのあとにE/iマークを付ける。これは、すべてのデータについて行なわれる処理である。

次にそのデータが活用語であるか、すなわち品詞情報がE, +, M, N, -



(注) 図中の①②の記号は以下のものを示す。これらの場合、活用形処理は行なわれない。

① は、非活用語（名詞・連体詞・副詞・接続詞・助詞 etc）・助動詞および記号等。

② は、変則活用、および活用不明のもの。

のいずれかであるかを判定する。活用語でなければライトエリアのデータを出力データとしてそのまま出力ファイルに書き込み、そのデータの処理を終え、次のデータへ進む。

一方、活用語であれば、仮の代表見出し・代表読みがなを、真の代表見出し・代表読みがなに変換する。そのため一番初めに、稿末に示した活用語尾テーブル（→表Ⅱ）の中から該当する活用行の語尾テーブルをゴビ・エリアにストアする。活用行語尾テーブルのストアの仕方は次のようにする。まず、活用形情報と活用行情報の2ケタを語尾テーブルのはじめの2ケタと比べ、等しくなければ次々とアドレスモディファイをして、等しいものを見つける。活用形・活用行情報と等しい記号をテーブル内に見いだしたら、そこから、次の右方のまでを、ゴビ・エリアの先頭から、その長さだけの間に移動させるのである。

つぎに、前にのべた、代表形見出し、代表形読みがなの、末尾の字の場所にそれぞれ a, a' というタグネームを、その一字前に b, b' というタグネームを与える。そうしてから、それぞれの代表形変換処理に進む。

(例)

驚	か	スペース	おどろか	スペース
b	a		b' a'	

(処理前の仮りの代表形見出し欄
代表形読みがな欄とタグネーム)

	来	スペース		こ	スペース
b	a		b'	a'	

(1) 五段・四段活用処理

活用形情報・活用行情報が、例えばFワだとすると、ゴビ・エリアにストアされた活用語尾テーブルでは

Fワネウワイエオッツネウわいえおっつ#

となっている。ここで、代表形見出しの、a にはいつている文字と、ゴビ・エリアの4字目(ウ)と比べ、同じでなければアドレスモディファイをしながら5字目、6字目……と順に比較して行く。そして、a の文字が、ゴビ・エリア内で等しい字を見つけたら、その字の左方に最初に見られる*印の右側の字を、a のところに移す。例えば、上の手続きにより、代表形見出し欄が「会エ」であれば「会ウ」に、「会っ」であれば「会ウ」というように、代表形に変

換される。

また、a にはいつている文字と等しいものがテーブルにない場合、すなわち順に比較して行っても同じ字が見つからず、印まで来た場合は、その#より左方に最初に来る*の左側の字を、a のところに移す。例えば、「会」「買」などは、上の手続きにより「会う」「買う」というように変換される。

次に代表形読みがな欄の処理をする。代表形読みがな欄の、a' にある文字を、#印から左方に最初に現われる*の右側の字に変えればよい。すなわち「かえ」「あお」等は、上の手続きでそれぞれ「かう」「かう」に変換される。

そして変換のおわった出力データを出力ファイルに書きこみ、次のデータ処理にはいる。

他の行の五段・四段活用語の処理も同様に行なう。

(2) 上一段・下一段活用処理

活用形情報、活用行情報が、たとえば「Gあ」であれば、ワークエリアには
Gあ*イルいる#

という活用語尾テーブルがストアされている。また、「Iら」であれば、ゴビ・エリアにストアされるテーブルは

Iあ*エルえる#

となる。

代表形見出し欄の、b にはいつている字が、ゴビ・エリアの4番目の字（イエなど）と等しければ、b と a にある2字を、それぞれゴビ・エリアの4・5番目の字にかえ、ゴビ・エリアの6番目の字（い、え等）と等しければ、b と a にある2字を、それぞれゴビ・エリアの6、7番目の字にかえる。例えば、「耐エレ」「通れれ」等は、それぞれ「耐エル」「通れる」に変換される。

上のどちらにもあてはまらなければ、a の字と、ゴビ・エリアの4番、6番目の字と比べる。a の字が、ゴビ・エリアの4字目または6字目と等しいときは、a の字の後（はじめはスペースがいつている）へ、ゴビ・エリアの5字目、7字目をそれぞれ入れてやる。この方法で例えば「耐エ」「通れ」等は、「耐エル」「通れる」に変換される。

以上の場合にはまだ当てはまらない、例えば「耐（ない）」「耐れ（ば）」等の処理は次のようにする。

まず、あらかじめ補助テーブルを下のようにとっておく。

＊ルレ＊るれ＃

この補助テーブルの2字目をはじめとして、(1)の五段・四段活用でやった処理と同様に、aの字が、ル・レであればそれを“ル”に、る・れであれば“る”に替える。またル・レ・る・れのいずれでもない場合は、aの文字のあとに“る”を加える。この方法で処理した「耐レ」「耐れ」「耐（ない）」等は、「耐ル」「耐る」「耐る」という形に変換される。これらは「耐エル」「耐える」等とは別見出しとなる。

代表形読みがなの変換処理は以下のようにする。すなわち、代表形読みがなのb'にある文字がゴビ・エリアの6字目と、等しければ、a'の文字をゴビ・エリアの7字目（る）に替え、等しくなければ、a'の後にゴビ・エリアの7字目の「る」を加える。この方法で「たえれ」「きれれ」「あえ」「しれ」等は、それぞれ「たえる」「きれる」「あたえる」「しれる」のように変換される。

そして処理の済んだ出力データを出力ファイルに書きこんで次のデータ処理にかかる。他の行の一段活用の活用語の変換処理も同様に行なう。

(3) 上二段・下二段活用処理

二段活用の場合、たとえば、活用情報が「Hあ」だとすると、ゴビ・エリアにHあ＊ウウ＃

という語尾テーブルがストアされている。

代表形見出しの、aにある字がゴビ・エリアの3字目、4字目と等しければそのままにしておく。また、aにある字が、(2)で出した補助テーブルの「ル・レ・る・れ」のどれかと等しければaの場所にスペースを埋めてその字を消す。また、上のいずれでもない場合はそのままにしておく。こうして処理すると「用ウレ」「用うる」「用る」「用」等は、「用ウ」「用う」「用」「用」というように変換される。他の行のものでも、「得る→得」「寝れ→寝」「起クル→起ク」のように代表形変換を行なう。

代表形読みがなは、a'の字がゴビ・エリアの5字目と、等しければそのまま

にしておき、等しくなければ a' に㊦を埋めて、そこにあった字を消す。

そして出力データを出力ファイルに書き込む。

他の行の二段活用の活用語も同様にして変換処理を行なう。

(4) 変格活用処理

たとえば活用情報が「Kか」とあれば、ゴビ・エリアの活用語尾テーブルは

Kか*クルクレコイコ㊦キ㊦*木*ク㊦くるくれこいこ㊦き㊦*米*く㊦#

となる。変換活用の処理はまず読みがなの方から行なう。代表形読みがなの、b'a' の2字と、ゴビ・エリアの3・4字目とを比べ、等しくなければ5・6字目、7・8字目……と順にくらべて、等しいものをさがす。等しい語尾が見つかった時は、そこから左の方で一番近い*の右側の2字をb'a'のところへ移して代表形にする。たとえば、この処理で、「こい」「くれ」はいずれも「くる」に変換される。等しい語尾が見つからない場合、すなわち、順々に比較して行って、ゴビ・エリアの比べるものが#になった場合は、代表読みがなのa'と次の一字分(スペース㊦がはいっている)とゴビ・エリアの3・4字目、5・6字目、7・8字目……と比べて行く。a'・㊦と等しいものをゴビ・エリア内で見つけたら、その等しい所から左方で一番近い*の右側の2字をa'・のところに移す。そうすることにより「こ(+㊦)」「き(+㊦)」はいずれも「くる」に、「く(+㊦)」は「く」という代表形で示されるようになる。

つぎに代表形見出しの変換を行なう。

代表形見出しの変換も、まず b・a の二字、または a とそのつぎの一字分の㊦とゴビ・エリアとを比較し、代表形読みがなと同じように処理する。そうするとテーブルにある語尾の「コイ」「キ(+㊦)」「クレ」等はすべて「クル」に「くれ」「こい」「こ(+㊦)」等はすべて「くる」に、「く」「ク」は「く」「ク」というように代表形に変換される。

上で処理できない場合、すなわちテーブルにある語尾を持たない形で書かれている場合は、まず、a とゴビ・エリアの4番目、6番目、8番目……と順に比べて行き、等しいものを探す。等しいものがみつかったら、そこから左方で一番近い*の右側の二番目の字を a に入れる。この処理で「来レ」「来い」は、それぞれ「来ル」「来る」に変換される。

それでもまだ処理できない、すなわち、順に比べて行っても等しいものがみつからずに、ゴビ・エリアの#まで来た場合は、その見出しは漢字である。そして、その場合、先に処理した読みがなの末尾（a'またはその後）が「る」であれば、代表形見出しのaのつぎに「る」を移し、そうでなければ代表形見出しをそのままにしておく。上の処理により、「来(き)」「来(こ)」「来(くる)」等は「来る」；「来(く)」は「来」という代表形に変換される。

以上の処理が済んだら、出力データを出力ファイルに書きこんで、次のデータへ進む。

他の行の変格活用の活用語の処理も同じようにする。

(5) 形容詞活用処理

形容詞型活用語（品詞情報はM, N, -）の場合にゴビ・エリアにストアされたテーブルは、

ⓂⓂ木イⓂクⓂウⓂカラカロカッカツケレカレウ木木いⓂくⓂうⓂから
からからかつかれゅう木木シⓂカリカルキⓂカレ木木しⓂかりかるきⓂ
かれ#

となっている。

このテーブルをもとに、(4)の変格活用と同じ方法で処理する。そうすると代表形見出しの、「美しく」「寒う」「恋しけれ」「早く」「寒れ」等であるものはそれぞれ「美しい」「寒い」「恋しい」「早い」「寒」に、代表形読みがなの、「うつくしく」「さむう」「こいしけれ」「はやく」「さむき」等であるものはそれぞれ「うつくしい」「さむい」「こいしい」「はやい」「さむし」に変換される。しかし、この方法では、代表形見出しの「美しき」「同じき」「美シキ」「淋しかり」等や、代表形読みがなの「うつくしき」「おなじき」「さびしかり」等が、それぞれ「美しし」「同じし」「美シシ」「淋しし」「うつくしし」「おなじし」「さびしし」等になってしまう。したがって一旦処理した代表形見出し・代表形読みがなの、末尾2字が「しし」「じし」「シシ」「ジシ」である場合だけうしろの「し」または「シ」の部分にⓂを埋めて、その字を消すようにしておく。

こうして処理した出力データを出力ファイルに書きこんで次のデータ処理へ

うつる。

4・おわりに

以上述べた理論で、動詞・動詞性接辞・形容詞・形容詞性接辞の、ほとんどすべての場合、代表形に変換することができる。上の理論をもとにして、(1)～(5)まで、それぞれを分けてプログラムを作り、テストを行なっているが、その結果、現在のところでは(1)および(2)ではすでに変換処理のできるようになっている。このあと、(3)、(4)、(6)が完成したら1本につなげて、実用的なプログラムになるわけである。

なお、その後、助動詞処理をすることのできるプログラムを作成することおよびエラーデータのチェック法の開発等の問題を残していることを付言しておく。

Fワネウワイエオツツネうわいえおっつ#
Fか*クカキケコイツツ*くかきけこいつ#
Fが*グガギゲゴイ*ぐがぎげごい#
Fさ*スサシセソネ*さしせそ#
Fた*ツタチテトツ*つたちてとつ#
Fな*ヌナニネノン*ぬなにねのん#
Fは*フハヒヘホ*ふはひへほ#
Fば*ブバビベボン*ぶばびべほん#
Fま*ムマミメモ*むまみめもん#
Fら*ルラリレロイツツ*るりれるいつ#
Gあ*イルいる# Gが*ギルきる#
Gか*キルきる# Gざ*ジルじる#
Gた*チルちる# Gだ*チルちる#

G な*ニルにる#

G は*ヒルひる#

G ば*ビルびる#

G ま*ミルみる#

G ら*リルりる#

G わ*キルるる#

I あ*エルえる#

I か*ケルける#

I が*ゲルげる#

I ぎ*ゼルぜる#

I た*テルてる#

I だ*デルでる#

I な*ネルねる#

I は*ヘルへる#

I ば*ベルべる#

I ま*メルめる#

I ら*レルれる#

I わ*エルゑる#

H あ*ウウ#

H か*クク#

H が*ググ#

H た*ツツ#

H だ*ツツ#

H は*フふ#

H ば*ブぶ#

H ま*ムむ#

H や*ユユ#

H ら*ルル#

J あ*ウウ#

J か*クク#

J が*ググ#

J か*スス#

J ぎ*ズズ#

J た*ツツ#

J だ*ツツ#

J な*ヌぬ#

J は*フふ#

J ば*ブぶ#

J ま*ムむ#

J や*ユユ#

J ら*ルル#

K か*クルクレコイコ㊦キ㊦*ク㊦*くるくれこいこ㊦き㊦*く㊦##

K さ*スルスレシロセヨセイ㊦サ㊦シ㊦*ス㊦*するすれしろせよせい㊦さ㊦し㊦*す㊦##

K ぎ*ズルズレゼ㊦*ズ㊦*ずるずれぜ㊦*ず㊦##

K な*ヌ㊦ヌル*ぬ㊦める##

㊦㊦*イ㊦㊦ウ㊦カラカロカッカツケレカレウ*い㊦く㊦う㊦からから
かっかつかれかれゅう*シ㊦カリカルキ㊦カレ*し㊦かりかるき㊦かれ#
(実際には、これらはすべて連続した形でテーブルになっている)