

# 漢字の読みの分類とその統計的性質

稲永 紘之 · 高木 康夫 · 吉田 将  
 (九州芸工大) (日本ユニバツ) (九大・工)

## 1. まえがき

我々は十数年来、カナ漢字変換を中心とする日本語機械処理のための辞書の充実に努めてきたが、辞書項目の登録には人手の介入が避けられず、それ起因した誤りの修正作業が辞書の大型化とともに大きな障害となってきた。

辞書の見出し部や漢字表記部の誤りを迅速に検出することを主眼において、JISの第1水準の全ての漢字を対象にして、我々は漢字の読み辞書を作成したが、この辞書では音韻変化形を含めた全ての読みを登録することとし、各読みには、常用漢字表に示されている読みを参考にして音訓の許容度のレベル付けや音韻変化の別によるパターン付けを行い、辞書内容のチェック以外の使用にも備えたものとした。

今回更に音読みについては、呉音、漢音、呉漢音(呉音と漢音が同じ音)、唐宋音、慣用音、国字音に分類してマーク付けを行い、チェックの終わった自立語辞書<sup>(1)</sup>を材料にして、日本語の単語における語構成の実態を漢字の読みの面から定量的に調べた上で、漢字の読みについての調査と併せて報告する。なお、音読みの分類は、角川書店版の「漢字の読み方(武部良明)」と「角川中辞典(貝塚他)」に拠った。

## 2. 漢字の読み辞書

### 2.1 漢字の読み

漢字の読みは大きく分けると、音読みと訓読みになるが、字によつては複数の音、訓を持ち、更に前後の文字により音韻の変化が生じ複雑な様相を呈している。また、音読みを歴史的にみれば、中国六朝期の南方の音が主体といわれる呉音、次いで隋唐時代の北方の音である漢音、更に11世紀以降の江南の音である唐宋音、それに誤読の慣用化による慣用音、数は少ないが、和製漢字の音に似せた発音である国字音があり、音読みの複雑さの原因となっている。一方訓読みは、もともと漢字の意味をもとに和語を当てたものであつたが、音韻変化、送り仮名のゆゑ、漢字との対応が不明確な当て字読みである熟字訓などにより、これも複雑と極めている。

### 2.2 読み辞書の作成

読み辞書の作成は自立語辞書<sup>(1)</sup>から自動的に行うこととし、8.6万語の自立語の表記に用いられている漢字に対して、その読みを辞書中で用いられている頻度とともに抜き出した。その作業の前に、自立語辞書はJISの漢字の判定を機に、出まゝ限りJISの第1水準の漢字を用いて書き改め、手作

常用漢字		教育漢字(配当学年)		カナ見出しに含まれる小文字数							
JIS漢字コード	沖漢字コード	音訓	音韻変化パターン	音訓	音韻変化パターン						
3555	1270	15	*101*カク	20	カカ	*342*カツ	*50	ケキ			
3556	4644	1	*101*ク	*50	ク	20	カ				
3557	2204	15	*101*ク	20	ク	*10	ク	42	ヒツ		
3558			60	ク	*501*ク	62	ク	60	カタ		
3559	0305	11	*101*ク	20	ク				62	カタ	
355A	4411	1	*101*ク	20	ク						
355B	7513	16	20	ス	@101*ク	41	ス				
355C	2411	13	20	ミ	.10	ク	*101*ク	.50	ク	*10	ク

表-1 漢字の読み辞書の一部

業でチェックを行った。しかし、使用頻度が低いとして出た読みで、音韻変化形などは、漢和辞典を引いても本当に正しいのかどうかわからず苦痛した。そこで、怪しげな読みをわざと落としたり読み辞書を用いて、再度チェックを行い、エラーとして出た語を詳しく調べるところを繰り返し読み辞書を修正した。また辞書に現れない漢字や読みについても漢和辞典から収録した。

2.2.1 読み分類

各読みに対して、表-1に示すように3つの観点から見た分類記号をつけた。

1行目は、音読みの漢、吳音などの種別によるもので、各記号の意味は次の通りである。

- “羊”：吳音
- “@”：吳漢音（吳音と漢音が同一）
- “\*”：漢音
- “.”：慣用音
- “+”：唐宋音
- “:”：国字音
- “|”：訓

2行目は、常用漢字表を参考にした音訓別の使用のための許容度のレベル分けをしており、各記号の意味は次の通りである。

- “1”：常用漢字表本表音訓欄に掲げられた標準の音読み
- “2”：同じく標準の訓読み

“3”：本表備考欄に掲げられた例を参考に、許容されると考えられる音韻変化した音読み

“4”：同じく音韻変化した訓読み

“5”：許容されはしないが一般に使われることもある音読み

“6”：同じく許容範囲外の訓読み

“7”：当2字読み

3行目は、音韻変化のタイプ分けをしたもので、各記号の意味は次の通りである。

“0”：音韻変化のない標準タイプ

“1”：標準タイプの読み末尾に送り仮名が付加されたタイプ

“2”：濁音あるいは半濁音化が起きたタイプ

“3”：“1”と“2”が複合して起きたタイプ

“4”：促音化、連声が起きたタイプ

“5”：“1”と“4”が複合して起きたタイプ

“6”：“2”と“4”が複合して起きたタイプ

表-2は、JISの第1水準の漢字に対して、上記の2行目と3行目の記号による2分類した読み分類表である。表中、括弧の中の数字は、読み辞書の音韻変化のタイプが“4”となつていふもののみを拾った場合の結果であり、その左横の数字は、辞書にはその容量節約のためにタイプ“0”に含

音韻変化のタイプ

音訓別許容度レベル	音韻変化のタイプ							計
	0	1	2	3	4	5	6	
1	2190	0	0	0	106( 6)	0	0	2296(2196)
2	1410	23	1	0	29( 25)	0	0	1463(1459)
3	0	0	423	0	172(172)	0	6	601( 601)
4	0	172	474	57	30( 30)	1	13	747( 747)
5	1458	0	116	0	151(113)	0	2	1727(1689)
6	2402	81	345	11	136(135)	8	12	2995(2994)
計	7460	276	1359	68	624(481)	9	33	9829(9686)

表-2 JIS第1水準漢字2965字の読み分類 (当2字読みを除く)

まれているが、本来“フ”にも含まれるべきものを“フ”に移した結果である。例えば、“カツ(滑)”という読み方は、“円滑”の中では“ウ”のタイプで、“滑草”の中では“フ”のタイプとなるが、辞書には、読み名に小文字が一つ含まれる可能性があるという情報を小文字数の欄に含ませ、“ウ”の記号のみを付けている。

表-2で、許容度のレベル記号が“フ”より上は、常用漢字のみについての、常用漢字表で許容されると考えられる読み名の統計表といえる。

表-2によれば、JIS漢字と常用漢字表で許容されている常用漢字の1文字当たりの平均の読み数は、次の通りとなる。

JIS漢字・・・3.32 (9829/2965)

常用漢字・・・2.63 (5107/1945)

ちなみに、許容外の読み名で含めた常用漢字の読みは3.81 (7405/1945)であり、常用漢字表では大幅な読み名の制限を行ってのこととわかる。

図-1は、JIS漢字の読み名に占める音読みと訓読みの比率を示したものであり、音読みについては、更に漢・呉音などの種別に細分類して、その比率を示したものである。

図-2は、常用漢字の許容されている読み名について、上と同様のことを行ったものである。

図-2の訓が図-1の訓に比べて比率が小さいのは、常用漢字表では和語の動詞などを記すための訓を大幅に制限していることを反映している。

### 2.2.2 読みの特徴など

当2字読みを除いた漢字の読み名の異なり総数は2387である。

音読み名の異なり総数は459であり、音読みでは、“フ”、“ン”、“レ”、“ペ”以外の仮名1文字で表わされる漢字が存在する。

仮名2文字以上で表わされる音読み

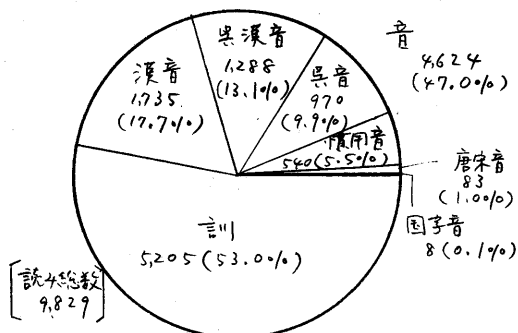


図-1 JIS漢字の読み

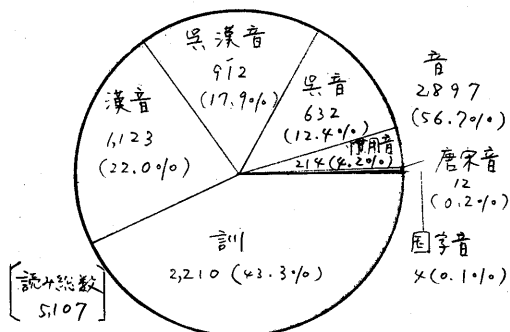


図-2 常用漢字(許容内)の読み

を末尾の仮名文字に8、2分類すると次の18種に限られる。度数と共に示せば

- |        |   |
|--------|---|
| ① “～ウ” | 1047  |
| ② “～ン” | 842   |
| ③ “～イ” | 493   |
| ④ “～ツ” | 376   |
| ⑤ “～フ” | 361   |
| ⑥ “～キ” | 75  |
| ⑦ “～ヨ” | 70  |
| ⑧ “～ユ” | 50  |
| ⑨ “～ヤ” | 41  |
| ⑩ “～チ” | 28  |
| ⑪ “～エ” | 3 (貝 <sup>へ</sup> , 平 <sup>へ</sup> , 平 <sup>へ</sup> ) |
| ⑫ “～カ” | 1 (博 <sup>か</sup> )                                   |
| ⑫ “～ナ” | 1 (健 <sup>な</sup> )                                   |
| ⑫ “～ノ” | 1 (信 <sup>の</sup> )                                   |
| ⑫ “～フ” | 1 (雑 <sup>ふ</sup> )                                   |
| ⑫ “～ミ” | 1 (心 <sup>み</sup> )                                   |
| ⑫ “～ゴ” | 1 (双 <sup>ご</sup> )                                   |
| ⑫ “～ブ” | 1 (三 <sup>ぶ</sup> )                                   |

11位以下は特殊なきで読みに近い読みであり、現代語においては上位の10種に限らざるをえてよさそうである。

音読みで、度数の多い10位までを挙げると、次の通りである。

- |           |           |
|-----------|-----------|
| ①“コウ” 102 | ⑥“カン” 59  |
| ②“ショウ” 94 | ⑦“ジヨウ” 57 |
| ③“シ” 76   | ⑧“キョウ” 55 |
| ④“トウ” 61  | ⑨“キ” 55   |
| ⑤“ソウ” 60  | ⑩“カ” 47   |

訓読みのみなり総数は2216であり、音読みのをれに比べると8.2倍に達する。

訓読みで、度数の多い10位までを挙げると、次の通りである。

- |         |         |
|---------|---------|
| ①“カ” 45 | ⑥“ハ” 37 |
| ②“オ” 44 | ⑦“ガ” 34 |
| ③“ト” 41 | ⑧“ス” 34 |
| ④“イ” 38 | ⑨“ヨ” 34 |
| ⑤“ス” 15 | ⑩“下” 33 |

訓読みで度数の多い読みは1文字の読みであり、2文字の読みは、36位に、度数14の“マ”と“カタ”がやると現われる。

訓読みにおけるの思なり数は、送り仮名の付け方によつて大幅に変るとも考えられる。我々の辞書では、送り仮名は極力落とすという方針を採つていさうだ。読みは短かいものが多い。

訓読みについて末尾の仮名で分類17カにわけて、特に興味ある結果は得られなかった。

この節での漢字の読み調査では、“ツ”、“ヤ”、“ユ”、“ヨ”の大文字と小文字の区別はしていない。

### 3. 自立語辞書中の漢字とよの読み

チェック済みの自立語辞書(86426語)を対象に、2.で述べた漢字の読み辞書を用いて、個々の読みを頻度と語頭、語中、語末の読みを三つに分けて取った。漢字1字のみで構成されている語については、促音化されるか変

動詞の語幹の読みのみ語頭の読みとし、それ以外は、語尾の読みとした。

また、2.2.1で述べた漢、吳音などの自立語辞書中での頻度も併せて取った。

なお、この実験のときには、読みの小文字数には留意してはなかつたので、読みの数などは、表-1における括弧の中のものを探ることとする。

読み辞書に登録していた読み数である。9688のうち、実際に自立語の春記に用いられていたのは、7900であったが、その音訓などによる構成の比率は、図-3に示す通りである。

また、読みを述べ数を用いた構成の比率は、図-4に示す通りである。

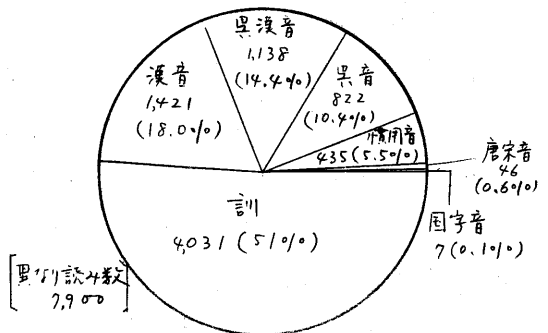


図-3 自立語辞書中での読み(異なり)

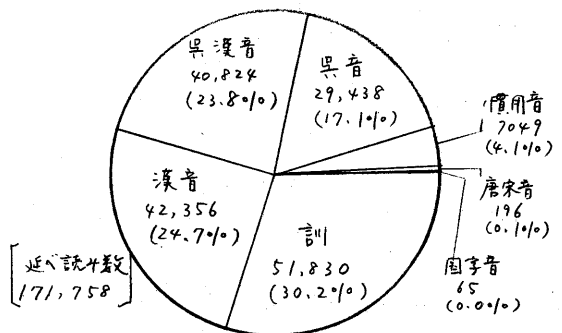


図-4 自立語辞書中での読み(延べ)

図-3と比べると図-4では訓の占める比率が著しく低下して、漢、吳漢、吳音の伸びが目立つ。これは、自立語の中では名詞が圧倒的に多く、名詞の

中では、漢語がまた圧倒的に多いことの現われと見ることが出来る。

読み辞書に登録されているが、自立語辞書中の漢字の読み方の調査に用いられたのは、表-3に示す158文字である。このうち、“絢”と“厩”と“猷”については、自立語辞書作成のミスで、“絢らん”、“厩舎”、“猷”がそれぞれ平仮名で表記されていたために用いられた。また、“囀”と“蒿”と“柘”は、自立語辞書に用いられているものの、それぞれ“牡囀”、“蒿麦”、“柘植”などにおいて、振り仮名を振ってないために、用いられた。また、自立語辞書中には、“竜”の異体字である“龍”もろくに用いており、これを“竜”に統一することにすれば、今後とも自立語辞書で用いられることとなるような文字は、153字ということになる。

娃始穉芦姐絢壹雄耐厩  
 新徽禦亨兕匡蕃莠杆杓  
 菟鞣夙舜渚蒲湘蔞奈訊  
 銚柘囀旋禎鄭菟瓠官鳩  
 籟疋彪彬斌冨鋪甫輔峯  
 劉亮琳麟吝澁豈亘

劃權繼梳栝苜侃恒湖館  
 良嵯遂塙塙確哂卓突勺  
 柁驛岱瀾異數物玟註瀟  
 塗盍傑剗豁馨晶醜構采  
 奎也曠繡瀟猷祐培慾裡

表-3  
 自立語辞書チェックの際用いられた漢字

4. 自立語の語構成のパターン化

日本語の単語における語構成の実際を、漢字の読み方の面から定量的に調べたために、まず自立語辞書から、外来語と、外来語交じりの語を取り去り、残りの81,047語を対象として、読み辞書の漢、吳音などの種別を基とする情報を利用して、自立語のパターン化を行った。

また、派生語を生み出す重要な要素である接尾語についてもパターン化を行い、その特徴を調べた。

ㄐ10 ㄐㄣ	0000ㄐ	20 ㄐ
ㄐ10 ㄐㄜ	*10 ㄐㄜ	42 ㄐ
*101ㄐㄜㄜ	*10 ㄐㄜ	42 ㄐ
ㄐ101ㄐㄜㄜ		
ㄐ101ㄐㄜㄜ		
*101ㄐㄜ	0000ㄜ	20 ㄜ
ㄐ10 ㄐㄜ	*10 ㄐㄜ	42 ㄜ
*10 ㄐㄜ	0000ㄜ	20 ㄜ

表-4 漢字の読み方からみた自立語の語構成のパターン化

表-4においては、“論じ合(う)”, “老巧搦(る)”, “良妻搦(る)”は、それぞれ、“@0ㄣ”, “@木ㄣ”, “\*羊ㄣ”とパターン化されることになる。ここで、“0”は、仮名文字列であることを示している。他の記号の意味は、2.1.1で述べた通りである。

4.1 品詞活用別パターン分布

品詞別、活用別にパターン分布を調べたが、そのうち特徴が著しいものを図-5, 図-6, 図-7に示す。

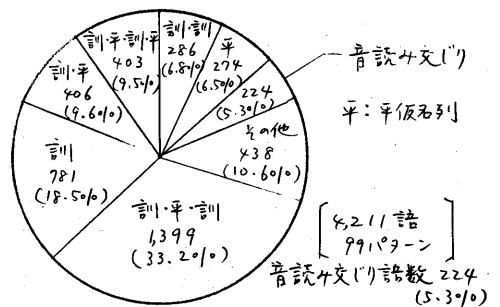


図-5 五段活用動詞パターン分布

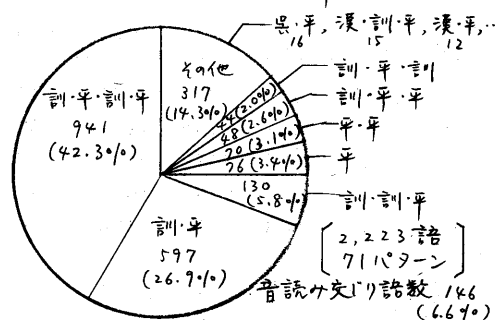


図-6 上一段下一段動詞パターン分布

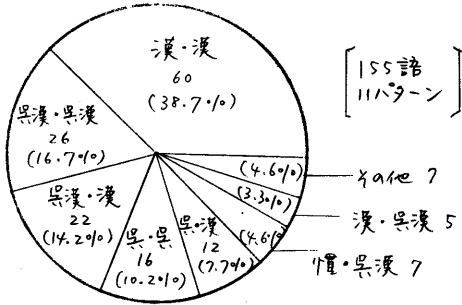


図-7 形容動詞タリ活用955漢字2文字パターン分布

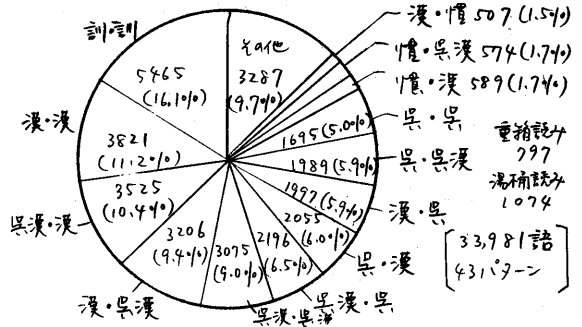


図-8 漢字2文字で表わされる自立語のパターン分布

図-5, 図-6 はいずれも和語の動詞が大部分を占めるので、音読み漢字が現れる比率は小さい。上一段・下一段動詞は自立語辞書中には不変化部命ということ、語尾が一部付けられた見出しとなつてゐるので、訓・平・訓・平というパターンに属する語が多い。

図-7での著しい特徴は、当然といえば当然だが、訓読みを含んだパターンが一つも存在しないことである。また、タリ活用の形容動詞はむづかしい漢字を用いて表記することが多いので、JESの第1水準の漢字だけでは、実に全部で272語のうち、107語が用ゐられ得ないのである。

4.2 漢字など2文字組のパターン分布

表-5と図-8に漢字など2文字組のパターン分布を示す。

前後	訓	漢音	呉漢音	呉音	慣用音	国字音	唐音	平仮名列
訓	5465	377	247	401	42	4	3	2349
漢音	198	3821	3206	1997	507	2	9	254
呉漢音	242	3525	3075	2196	473	3	12	245
呉音	306	2055	1989	1695	354	3	8	149
慣用音	49	589	574	364	117	4	2	37
国字音	0	1	0	2	42	0	0	1
唐音	2	2	10	0	1	0	7	1
平仮名列	443	280	199	180	25	0	1	863

表-5 漢字など2文字組みのパターン分布表

4.3 接尾語のパターン分布

図-9に接尾語3469のパターン分布を示す。名詞、動詞の連用形、変動詞語幹のそれぞれに付く接尾語についてのパターン分布も求めていくが紙面の都合で省略する。

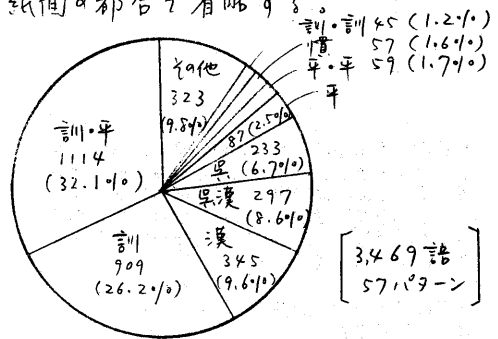


図-9 接尾語のパターン分布

5. あとがき

もともと辞書内容のチェックのため始めた仕事であったが、自立語の語構成の複雑さを目の当たりにするようになった。漢音、呉音の認定は、漢和辞典により異なるので、現状では、どれに従うかで結果も少し異なる。

[文献]

[1] 福永吉田:「日本語の理のための機械辞書」情報処理, 23-2