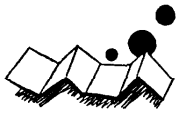


解説



日本語処理のための機械辞書†

稲永 紘之†† 吉田 将†††

1. ま え が き

かな漢字変換装置をはじめとする日本語処理装置の発達は、ここ数年目覚ましいものがある。社会の国際化・情報化に伴い、機械（支援）翻訳、日本語による情報検索などの要求は強まる一方であるが、これらを実現するためには日本語処理のための本格的な機械辞書が必要不可欠である。我々は十数年前、カナ漢字変換を中心課題として機械辞書の作成を開始し¹⁾、以来、より高度で多様な日本語処理にも対応できるよう辞書内容の充実に努めてきた²⁾⁻⁴⁾。すなわち、仮名書きあるいは漢字仮名交じり書き日本語文の形態素・構文・意味解析、未登録語・派生語・同形異義語処理、言語情報を利用する音声・文字認識などを念頭に置きながら、必要で効果的と思われる情報は可能な限り取り入れた。辞書の増補・改訂・校正の作業は常時進められているが、8万7千余の基本語（短単位の自立語）については、校正も完了したので研究利用に限り公開使用に供することにした。

本辞書は大別すると、自立語に関する辞書と付属語に関する辞書からなり、以下のような辞書群から構成されている。

- | | |
|--------------------|----------|
| ① 基本語辞書 | } 自立語原辞書 |
| ② 複合語・係り受けの句・慣用句辞書 | |
| ③ 対義語辞書 | |
| ④ 接頭・接尾語辞書 | |
| ⑤ 漢字の読み辞書 | } 付属語原辞書 |
| ⑥ 助動詞辞書 | |
| ⑦ 助詞辞書 | |

本文では、これらの辞書 - 自立語原辞書と付属語原辞書 - の各項目を示し、その項目を辞書に収容することの必要性、問題点等を解説する。この辞書を実際の

処理に使用する場合には目的に応じて適切な記憶構造に再編成し、検索システムを用意する必要がある。本文ではこの点については触れないが、長尾⁵⁾、和智⁶⁾等の報告がある。

2. 自立語原辞書

2.1 見出し語句の登録

① 各種の国語辞典、用字用語辞典、アクセント辞典、新聞・雑誌記事、文芸作品などから、広く現代社会で用いられている一般の語と句を収集し登録した。

② 登録語数は、昭和56年10月現在で基本語（派生語も含めた短単位の自立語）⁷⁾ が8万7千余、複合語（長単位の自立語で、基本語の接続した語）・係り受けの句・慣用句が12万5千余、対義語の組が2千3百余である。

接頭・接尾語辞書及び漢字の読み辞書は、本辞書の他の内容から自動生成される（3.2 (U) 及び2.2 (F) 参照）ので原辞書としては作成していない。

2.2 自立語原辞書の構成

自立語原辞書は、図-1のような項目からなる。各部の内容は以下の通りである。

(A) 見出し語句 ID 部 (5桁)

英数字からなる5文字列で、その語句の登録された順番を示し、校正の際に利用される。

(B) 校正済み表示部 (1桁)

(C) カナ見出し部 (18桁)

カナ大文字 (46字。ただし、'ワ' は '#' で代用)、濁点、半濁点、長音記号 (マイナス記号で代用)、語や句の切れ目記号 '†'、'*' の計51文字を用いて表わした。(一部3.2で説明)

① 活用語はその語幹で示したが、上一段・下一段動詞については語尾の一部も含めたいわゆる不変化部分で示した。基本語としての'来る'、'する'は語幹がないので登録されていない。

② 外来語は原則として原語のワードの切れ目に切れ目記号 '†' を入れて示したが、原語がハイフンで

† Machine Dictionary for Japanese Language Processing by Hiroyuki INANAGA (Kyushu Institute of Design) and Sho YOSHIDA (Kyushu University).

†† 九州芸術工科大学

††† 九州大学工学部

見出し ID		カナ見出し			見出し切れ目		文節切れ目		文法・意味情報											漢字	表記				
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

- | | | | |
|----------------|----------------|------------------|--------------|
| A: 見出し語句 ID 部 | H: 含自立語の表示部 | O: 活用の行表示部 | U: 語句構成表示部 |
| B: 校正ずみ表示部 | I: 小文字数表示部 | P: 形容詞の種類表示部 | V: 普通名詞部 |
| C: カナ見出し部 | J: 同形異義語句優先順位部 | Q: 形容動詞の種類表示部 | W: 代表漢字数部 |
| D: 格助詞部 | K: 重要語表示部 | R1: 副詞の種類表示部 | X: 固有名詞部 |
| E: 古語・外来語表示部 | L: 自・他動詞種類表示部 | R2: 副詞に後続可能文字表示部 | Y: 助数詞, 代名詞部 |
| F: 見出し語句切れ目表示部 | M: サ変動詞表示部 | S: 接続・感動・連体詞部 | Z: 漢字表記部 |
| G: 文節切れ目表示部 | N: 活用の種類表示部 | T: 見出し語句の種類表示部 | |

図-1 見出し語句の構成

結ばれている場合のように一語と考えられる可能性のある語については、切れる可能性のある箇所に切れ目記号「*」を入れて示した。

‘アト↑ランダム’, ‘スピード*アップ’

③ 複合語は、それを構成する基本語と基本語の切れ目、あるいは分かち書きをした場合には切れる可能性のある箇所に切れ目記号「*」を入れて示した。

‘キョウサン*シユキ’ [共産主義], ‘ウカヌ*カオ’ [浮かぬ顔]

(D) 格助詞部 (2桁) 3.2 で説明。

(E) 古語・外来語表示部 (1桁)

古語, 外来語あるいは外来語を含む語句は表示を入れて表わした。

(F) 見出し語句切れ目表示部 (8桁)

カナ見出し部 (C) と漢字表記部 (Z) を対応づけ、漢字の読みを与えるもので、カナ見出しのどのカナ文字列がどの漢字に対応するかを示した。一つの切れ目は数字1桁で表わした。これから漢字一字の読み辞書を自動生成することができる⁹⁾。

ケイサンキ 2|2|1 計算機

カナ見出し 切れ目 漢字表記

(G) 文節切れ目表示部 (8桁)

カナ見出し部の文字数あるいはカナ見出し部に切れ目記号「†’, ‘*’ が含まれるときは、その記号間の文字数を2桁の数字の組を最大4組使って表わした。カナ入力文とカナ見出しのマッチングの際使用する情報である。

(H) 自立語を含む自立語の表示部 (1桁)

次の二つの条件のうち少なくとも一方を満たす自立語に対して表示を入れた。

① その自立語だけで文節となり得るが、それより短い他の自立語に付属語 (列) が付いた文節と、カナ表記が一致する。

‘ホンカ’ [本科] は、‘ホン’ [本] に係助詞 ‘カ’ が

付いた ‘本カ’ とカナ表記が一致する。

② その自立語に何らかの付属語 (列) を付け文節を作ったとき、その自立語より短い他の自立語に付属語 (列) が付いた文節とカナ表記が一致する。

‘クルマ’ [車] に格助詞 ‘デ’ が付いた ‘クルマデ’ は、‘クル’ [来る] に副助詞 ‘マデ’ が付いた ‘クルマデ’ とカナ表記が一致する。

(I) 小文字数表示部 (1桁)

カナ見出しに本来含まれるべき小文字 (‘イシャ’ の ‘ゃ’ など) の数を数字1桁で表わした。

(J) 同形異義語句の使用優先順位部 (1桁)

カナ表記は一致するが、漢字表記が異なる語 (句) を、互いに同形異義語 (句) という。これらの語句に対して、あらかじめ使用頻度を推定して優先順位を付した。

(K) 重要語表示部

国立国語研究所「分類語彙表」の重要語を参考に、一部我々の判断で追加・削除したもの。

(L) 自・他動詞の別、格の種類表示部 (2桁)

基本語動詞についてはそれが他動詞である場合には他動詞部 (1桁) に表示した。その他の場合については3.2で述べる。

(M) サ変動詞の表示部 (1桁)

サ変動詞を、口語と文語、また ‘～スル’ と ‘～ズル’ のように清音と濁音で区別して、計4種類に分けて表わした。

(N) 活用の種類表示部 (1桁)

サ変以外の動詞を、①四段(五段)、②上一段、③下一段、④上二段、⑤下二段、⑥ナ変、⑦ラ変、⑧カ変、⑨四段と上二段、⑩四段と下二段、⑪‘～トスル’ のように活用する動詞 (‘生キ生キトスル’), に分類して表わした。

(O) 活用行の表示部 (1桁)

(N) の活用の種類に対してその活用行を示した。

表-1 名詞の分類

コード	内 容
1	他の34のいずれにも入り得ないもの(抽象的ものごとなど未分類のものを入れた)
2	時, 方角, 数量など, 副詞的用法のあるもの
3	'ナ'がついて, 形容動詞的にも使われることがあるもの
4	人間, 神, 仏を表わすもの
5	動物名, 動物を表わすもの
6	植物名, 植物の組織の一部を表わすもの
7	鉱物名, 鉱物, 自然物を表わすもの
8	飲食物, 料理を表わすもの
9	人間や動物の体, あるいは体の一部を表わすもの
A	衣服, 織物, 装身具など, 人間が身に着けるもの
B	建物, 橋など建造物を表わすもの
C	道具, 日用品, 簡単な機構の器具を表わすもの
D	書籍, 評論, 文章など読み書きの対象を表わすもの
E	乗物, 交通機関を表わすもの
F	病気, 怪我, 傷などを表わすもの
G	弾く楽器, 楽器一般を表わすもの
H	吹く楽器
I	打ったりたたいたりする楽器
J	動物, 植物, 人間も含めた広い意味での生物を表わすもの
K	貨幣, 金銭の他, 経済活動関係を表わすもの
L	建物の一部をなす部分, 部品を表わすもの
M	複雑な機構の機械などを表わすもの
N	自然現象を表わすもの
O	音楽, 曲の他, 芸能関係を表わすもの
P	期間, 時間を表わすもの
Q	衣服, 織物, 装身具などの一部を表わすもの
R	スポーツ, 遊びに用いられる道具, 用具を表わすもの
S	スポーツ, 遊びの他, スポーツ関係を表わすもの
T	時刻, 時期を表わすもの
U	人間が自然物に手を加えて作った道や壘などを表わすもの
V	建物の内部の部屋などを表わすもの
W	世界や活動の分野を表わす抽象的な場所を表わすもの
X	団体, 組織, 人の集まりなどを表わすもの
Y	光景, 様子, 状態を表わすもの
Z	物事が行れる具体的な場所を表わすもの

(P) 形容詞の種類表示部 (1桁)

形容詞を, ①口語, ②文語ク活用, ③文語シク活用, ④口語と文語ク活用, ⑤口語と文語シク活用に分類して表わした。

(Q) 形容動詞の種類表示部 (1桁)

形容動詞を, ①口語, ②口語で連体形に'～ノ'の形のあるもの, ③タリ活用, ④ナリ活用, ⑤口語とナリ活用に分類して表わした。

(R1) 副詞の種類表示部 (1桁)

副詞を, その働きから, 状態, 程度, 陳述, 及びこれら二つ以上を兼ねた副詞に分類して表わした。

(R2) 副詞の後続可能文字表示部 (1桁)

その副詞に'ト', 'ノ', 'ダ'が後続可能か否かを表わした。

(S) 接続詞, 感動詞, 連体詞表示部 (1桁)

接続詞, 感動詞, 連体詞, あるいはこれら二つ以上

を兼ねた語の別を表わした。

(T) 3.2 で説明

(U) 3.2 で説明

(V) 普通名詞部 (2桁)

普通名詞を表-1の意味分類コードを用いて35のクラスに大分類した。この名詞の意味は第二義までとした。(2.4 参照)

(W) 代表漢字数部 (1桁)

普通名詞の意味分類コードの第一義のものに対してその名詞の末尾から何個の漢字表記部の漢字などが意味の主体を成しているかを表わした。(2.4 参照)

(X) 固有名詞部 (1桁)

固有名詞を, ①地名・国名, ②人名・神仏の名前, ③建築・建造物の名前, ④団体・組織の名前, ⑤年号名, ⑥言語名, ⑦物語・書籍の名前, ⑧市名, ⑨県名に分類して表わした。

(Y) 助数詞, 代名詞表示部 (1桁)

助数詞, 人称代名詞, 指示代名詞の別を表わした。ここでいう助数詞には, 数字に続けて書かれる語であれば, 接尾語のみでなく, 広く名詞も含めた。

(Z) 漢字表記部 (32桁)

カナ見出し部(C)と切れ目表示部(F)に対応して, JIS 非漢字・第1水準漢字コードと, 特殊コード'000016'を用いた最長8個のコードよりなるコード列で表わした。'0000'は, それに対応するカナ文字列が, もともと平仮名か片仮名で表記されるべき文字列であるか, あるいは対応する漢字がJISの第1水準以外の漢字であることを示している。

2.3 対義語辞書

対義語辞書は, 関連語あるいは反対語が互いに関連のある文中に現われる傾向があることに注目して用意されたものである。たとえば, 'ショウシャ【勝者】'と'ハイシャ【敗者】'が対義語の組として登録されているとき, 一連の入力文中に名詞として'ショウシャ'と'ハイシャ'が現われたときには, それぞれの同形異義語群中から'勝者'と'敗者'を優先的に選ぶことができる。(辞書構成については3.2参照)

2.4 意味情報

意味情報は名詞に対して付加し, 2.2のように, まず, 固有名詞は9, 普通名詞は35のクラスに大分類した。さらに, 普通名詞に対しては, それが名詞性であるか, 動詞性であるか, 形容動詞性(述部で用いられることが多い)であるかを明らかにした。また, 名詞性と動詞性の名詞については, 末尾から何個の漢字表

記部の漢字などがその意味の主体を成すかを、その名詞を用いた文節が係りとなる名詞一名詞型、名詞一用言型の係り受けの句を調べることにより示した。(2.2の(W))これは、漢字情報を用いた名詞の細分類であるが、その類に含まれる名詞が実際に文中に現われる際の構文をも考慮に入れた分類であることに特長がある。

⑩: 末尾の漢字が意味の主体を成さない名詞性の名詞。‘殺人’、‘営業’、‘作曲’

①, ②, ③, ④, ⑤: それぞれ末尾の一, 二, 三, 四, 五個の漢字等が意味の主体を成すもの。‘恩人’, ‘異分子’など。④以上は慣用句にしか存在しない。⑥, ⑦, ⑧, ⑨は動詞性の名詞で、それぞれ末尾から一, 二, 三, 四個の漢字等が意味の主体を成すもの。‘転落死’, ‘山登り’など。

3. 係り受け関係辞書——意味解析への接近

3.1 係り受け関係

文は文中に現われる自立語が表わす概念の間にたまたま、ある関係(文節間の係り受け関係によって表わされる)が成り立っていることを主張するものである。したがって、文の意味を解析するにはまず文節間の係り受けの構造を求めることが重要である⁹⁾。係り受け関係の基本は2文節間の係り受け関係および動詞の格構造に分解して考えることができる。我々は2文節間の係り受けの関係を三つのレベルに分けて辞書に登録した。たとえば、‘列車が遅れたので試験を受けなかった’においては、‘列車が’は品詞レベルでは‘遅れる’と‘受ける’のいずれの文節にも係り得るが、意味上からは‘受ける’には係り得ない。しかし、‘列車が妨害を受ける’の場合は‘列車が’は‘受ける’に係るので2文節のみでは係るかどうかは決め難い。列車の遅れは乗物としての日常的な関心事の一つであり、列車が妨害を受けるのは何らかの目的で動作するものと一般としての列車であり、列車固有の日常的な事柄とはいえない。このような意味で、‘列車’と‘遅れる’は係り受けの関係があり(強い)‘列車’と‘受ける’は係り受けの関係がない(弱い)と考えたものを第1レベル(個々の語レベル)の係り受け関係と呼ぶことにする。我々は第1レベルの関係を係り受けの句あるいは係り受けの句に転用可能な複合語(‘列車*遅れ’は‘列車が遅れる’に転用可能)という形で可能な限り辞書に登録した。第2レベルは、漢字情報を用いた分類レベル(2.3の(W), 2.4, 3.2の(U)参照)に対応した

ものである。第3レベルは、表-1による名詞の大分類レベルに対応する係り受けの句である。たとえば、動物を示すクラス5に対して、‘5↑ナ [(動物) (が) 鳴く]’という係り受けの句を登録しておけば、‘5’のところに、5に属する名詞が代入可能であることにより、用例の大幅な減少を図ることができる。しかし、‘鳴く’ことのない動物もいるわけで、あいまいさが生じるのは避けられない。

係り受けの句や複合語を辞書に大量に登録することは構文解析およびそれに引き続く意味解析に重要な情報を提供するものと考えられ、さらに高度な意味解析には各語ごとの意味辞書が必要であり本辞書の範囲を越えている。

以下、係り受け関係・複合語に関連した辞書項目について説明する。

3.2 係り受け・複合語・対義語に関連した辞書

この辞書は自立語原辞書の中に含まれ図-1の辞書構成と同一のものである。

(C) カナ見出し部 (一部2.2で説明済み)

① 係り受けの句・慣用句は、それを構成する文節と文節の切れ目には切れ目記号‘↑’を、基本語と基本語の切れ目には切れ目記号‘*’を入れて示すことを原則とした。また、他の助詞や助詞列に変換可能な格助詞‘が’, ‘ヲ’, ‘ニ’, ‘ト’, ‘デ’, ‘カラ’(以下これらを標準格助詞という)を介した係り受けの句では、その格助詞を除いて、係りの基本語と受けの基本語の切れ目に切れ目記号‘↑’を入れて示した。

シラヌガ↑ホトケ [知らぬが仏] (慣用句)

メ↑アケ [目(を)開け(る)] (名詞一用言型の係り受けの句)

② 対義語の組とは、関連語、反対語などの基本語を二つあるいは三つ列挙したもので、基本語と基本語の切れ目に切れ目記号‘*’を入れて示した。

ショウシャ*ハイシャ [勝者・敗者]

アシ*テ [足・手]

(D) 格助詞部 (2桁) (一部2.2で説明済み)

係り受けの句のカナ見出し部には、(C)の①で述べたように標準格助詞は書かなかったが、その代りにここにその六つの格助詞をそれぞれ1~6の数字に対応させてコード化して表わした。

(L) 自・他動詞の別、格の種類表示部 (2桁) (一部2.2で説明済み)

基本語動詞、名詞一動詞型の係り受けの句については、その動詞が要求する格助詞((D)で与えられた以

外の)のうち基本的なものをつ、カナ見出しの最後の切れ目記号として‘*’を含み最後の基本語が動詞性名詞である複合語については、その複合語を係り受けの句に転用する際に、最後の‘*’の直前の名詞の後に用いられるべき標準となる格助詞をそれぞれ数字1桁で符号化して表わした。

符号化した格助詞あるいは格助詞の組を次に示す。

他動詞部: ‘ヲ’のみ

自動詞部: ‘ガ’, ‘ヲ’, ‘ニ’, ‘ト’, ‘ヲ’ と ‘ニ’ の組, ‘カラ’, ‘デ’

ゴ*コウセイ [語構成] (語を構成する)

(T) 見出し語句の種別表示部 (1桁)

カナ見出し部 (C) に切れ目記号 ‘†’ を少なくとも一つ含む語句に対しては、それが① (E) の表示と合わせて外来語, ②名詞一名詞型, 用言一名詞型, 名詞一用言型, 名詞一名詞一用言型の係り受けの句, ③慣用語, ④名詞一副詞一用言型の係り受けの句, ⑤副詞一名詞一用言型, ⑥連体形一名詞一用言型の係り受けの句, ⑦副詞一副詞型, 副詞一用言型の係り受けの句, であることを表わした。ここで、連体形には用言の連体形の他に名詞に ‘ノ’ に付いた連体修飾節も含み、副詞には副詞の他に、同形異義語が存在してかつ使用頻度の高い場合は用言 (主として形容詞, 形容動詞) の連用形も含んでいる。カナ見出し部に ‘*’ は含むが ‘†’ は含まない語句に対しては、① (E) の表示と合わせて外来語である。②対義語の組である、③最後の ‘*’ の直前の名詞の後に格助詞 ‘ノ’ あるいは標準格助詞をつけて係り受けの句としても使用できる複合語であるか否かを表わした。

(U) 語句構成表示部 (2桁)

① 基本語 (カナ見出し部に ‘†’ も ‘*’ も含まぬ語) に対しては、見出し語句切れ目表示部 (F) の末尾の切れ目表示に対応する接尾語 (いわゆる造語成分といわれる語も含む) が、一般的にその一つ前の切れ目表示に対応する漢字に後続して接尾語になり得るならば、その程度を3段階で表わし、同時にその漢字の意味分類を示すコード (V参照) も入れた。これらの情報は基本語辞書から自動的に接尾語辞書を作成する時に用いられる。たとえば、基本語が ‘コウチョウシツ’ [校長室] であったとすれば切れ目表示部は ‘232’ であり、末尾の切れ目表示 ‘2’ に対応する漢字 ‘室’ が一般的にその一つ前の切れ目表示部 ‘3’ に対応する漢字 ‘長’ に対してかなりの割合で後続するので程度は中程度とし、‘長’ の意味 (人間) を表わすコード

とともに示されることになる。

② 基本語以外の語句に対しては、係りの名詞部の末尾の漢字情報を用いて、名詞一名詞型, 名詞一用言型の係り受けの辞書作りが行えるなら、その程度を3段階で表わし、同時にその末尾の漢字の意味分類を示すコードも入れた。これらの情報は、係り受けの句あるいは係り受けの句としても使用可能な複合語から自動的に漢字情報を用いた係り受け辞書を作成する時に用いられる。たとえば、名詞全体の集合のうち末尾の漢字が ‘酒’ でありこれが意味の主体をなしている名詞のクラスを考えよう。このクラスには、‘甘酒’, ‘樽酒’, ‘合成酒’, ‘日本酒’, ‘密造酒’ など多くの ‘~酒’ が属する。このクラスに対して、‘酒†ヨ [(に) 酔 (う)]’, ‘酒†ノ [(を) 飲 (む)]’, ‘酒†アジ [(の) 味]’ のような係り受けの句を登録すれば、第一レベルの係り受けの句の直接登録に比べれば登録数の大幅な節約が図れる。しかも ‘~酒を飲む’, ‘~酒に酔う’, ‘~酒の味’ は ‘酒’ が意味の中心を成す名詞には普遍的に使われる表現と考えられるので程度は強いとして ‘酒’ とともに登録される。このとき、末尾の漢字 ‘酒’ の読みは ‘シュ’, ‘サケ’, ‘ザケ’ のいずれでもよい。また、この漢字情報を用いた係り受けの句 (第二レベルの係り受けの句) は、第一の単語レベルの係り受けの句のコーディングの際に、図-1のUの項にその情報を書き込んでおけば自動的に登録される。

4. 付属語辞書

付属語としては助動詞、助動詞が代表的なものである。また、‘なければならない’, ‘こともある’, ‘に対して’ などのように一まとまりで助動詞的、助詞的に使われ、きまった意味を有する表現 (連語) も (拡張された) 付属語の中間に入れて考えた方が便利な場合もある (意味処理を伴う場合など)。後者の、拡張された付属語まで含めた場合の処理については首藤¹⁰⁾の詳しい報告があるので、ここでは前者の助動詞、助詞のみを付属語とする場合の辞書について説明する¹¹⁾。文節は一般に自立部の後に0個以上の付属語が接続したものであるが、自立部・付属語間あるいは付属語相互間には明確な統語上の接続の規則がある。助動詞の接続順序については多くの資料があり、我々も接続順序表を作成し使用しているがその構造を述べる紙面がないので本文では原辞書の項目とその説明にとどめる。

4.1 助動詞原辞書

見出し語は助動詞29語 (活用形別では281種)、副

見出し	文節末	先頭	動詞判定	活用形	意味	接続	活用型	活用形	語幹、語尾	接続助動詞 1	接続助動詞 2
A	B	C	D	E	F	G	H	I	J	K ₁	K ₂

図-2 助動詞原辞書の項目

表-2 助動詞の活用形

コード	動詞型	形容詞型	形容動詞型	特別活用型 無変化型	備考
0 (注)					*1: 5段動詞
1	未	未	未		*2: カロ
2	志*1	志*2	志*3	志	*3: ダロ
3					*4: ク
4	用	用*4	用*5	用	*5: テ, ニ, ト
5	音*6	音*7	音*8		*6: イ音便 ウ音便 促音便
6	音*9	音*A		音	*7: カツ
7	終	終	終	終	*8: ダツ
8					*9: ン
9	体	体	体	体	*A: ウ, ユウ, シウ
A	仮		仮		
B	仮+バ	仮+バ		仮+バ	
C	命				
F				体+ヲ	

(注): 語幹、あるいは他の活用形に共通な先頭文字、無活用語も含む
 未: 未然形 志: 志向形 用: 連用形 音: 音便形
 終: 終止形 体: 連体形 仮: 仮定形 命: 命令形

詞 12 語、準体助詞 'ノ' の合計 42 語が登録されている。助動詞原辞書の項目を 図-2 に示す。

(A) 助動詞または助動詞に先行する助詞

大多数の助詞は助動詞の先行語とはならないが、準体助詞 'ノ'、副助詞 'マデ'、'ホド'、'クライ'、'ナド' など (12 語) は先行するので助動詞辞書の方に含めた。

(B) 文節末になり得るか否か

助動詞の未然形などは文節末になり得ない。

(C) 後続文節の先頭として要求される語

'イカザルヲ' という文節は 'エナイ' という後続文節を要求し、下一段動詞 'エル' を必要とする。

(D) 後続文節の動詞判定情報

後続文節として要求される動詞の文法情報。'ナラナクナル' (ラ行五段動詞を要求)

(E) 後続する助動詞の活用形

'タガル'、'～シカナイ' などに 'ダ' が接続する場合、その仮定形 'ナラ' しか後続しない。

(F) 見出し語の意味

受身、可能、自然、尊敬、使役、丁寧、過去、完了、推量、打消、様態、伝聞、希望、断定、比況の 15 種類のいずれであるか。

表-3 先行語の品詞と活用の型

品詞
5 段活用動詞、5 段活用動詞型助動詞
1 段活用動詞、1 段活用動詞型助動詞
カ変動詞、サ変動詞、ザ変動詞
形容詞、形容詞活用型助動詞
形容動詞、形容動詞活用型助動詞
特別活用型助動詞
無変化型助動詞
格助詞、準体助詞
接続助詞
副助詞

(G) 先行語の情報

未然形、連用形、終止形、連体形、体言、助詞、語幹のいずれに接続するか。

(H) 見出し語の活用型

動詞活用型、形容詞活用型、形容動詞活用型、特別活用型、無変化型のいずれであるか。

(I) 見出し語の活用形 (表-2 参照)

各活用形ごとに見出し語を設けた。

(J) 語幹または語尾

見出し語のどの部分が語幹 (語尾) であるかを示した。

(K₁), ..., (K_n) 後続する助動詞

後続可能な助動詞の全リストを示した。

4.2 助動詞原辞書

助動詞原辞書の項目を 図-3 に示す。

(A) 助詞見出し

助詞列 135 (うち単独の助詞 47 語 (終助詞は除いた)、助詞二つの列 86、助詞三つの列 'ヨリ・ホカ・ニ'、助詞四つの列 'ヨリ・ホカ・ニ・ハ') が見出しとして登録されている。

(B) 助詞の種類

格助詞、係助詞、副助詞、接続助詞、準体助詞、間投助詞のいずれか。

(C₁), ..., (C_n) 先行語の品詞 (表-3 参照)

(D₁), ..., (D_n) 先行語の活用型 (表-3 参照)

(E₁), ..., (E_n) 先行語の活用形

(F₁), ..., (F_n) 2 種以上マッチングの可能性のある場合の優先順位

助詞見出し	種類	先行語 1				先行語 2				見出しの切れ目
		品詞	活用の型	活用形	順位	品詞	活用の型	活用形	順位	
A	B	C ₁	D ₁	E ₁	F ₁	C ₂	D ₂	E ₂	F ₂	G

図-3 助詞原辞書の項目

助詞列で二つ以上の切り方がある場合や単独の助詞で2種以上の機能の違いがある場合の優先順位。

(G) 見出しの切れ目

見出しが助詞列である場合の切れ目(何文字目か)を示す。

5. あとがき

九州大学の故栗原俊彦教授を中心とした言語研究グループでは、20年余り前から多方面に亘る資料収集を行いつつ自然言語の機械処理の可能性を追求してきた。その一環としてカナ漢字変換を中心とした日本語処理の研究を始めたのは昭和38年であった。機械辞書の作成を開始したのは昭和41年からである。この辞書は昭和43年には大幅な改訂を行ったが、まだ見出し語句の選択方針、漢字表記の方法、送り仮名の不統一、同形異義語と派生語の処理能力などに多くの問題点が認められたので、昭和45年の秋から九大と九州芸工大稲永の協力の下に、それまでに蓄積された経験をもとに全面改訂に踏み切った。当初は、小型の国語辞典並みの見出し語は少なくとも必要だということで、岩波国語辞典(第2版)の見出し語をすべて登録したが、国語辞典の性格上、派生語(特に漢字3文字以上で表わされる語)はほとんど収録されていなかった。そこで基準を設けて追加登録を開始した。また、同形異義語の処理には文節相互の係り受け関係を利用することが有効であるとして用例(第1レベルの係り受けの句)の収集に力を注いだ。昭和54年秋から約1年半をかけて基本語の見直しを行い、JIS第1水準の漢字はすべて用いることとし、漢字を追加して修正した。また同時に、名詞の大分類の漏れを補い末尾の漢字による分類作業も基本語については完了した。用例の収集と辞書内容のコーディングは稲永が行い、辞書の保守は九州芸工大小西彬允氏の全面的な協力を得た。

本辞書の作成を始めたのは十数年前であり、当時の計算機システムのレベルに従って必要な情報を取り入れたため、現在では冗長と思われる情報、加えておけば良かった情報など不満な面は多々存在する。また、紙面の都合で本文では説明できなかった情報も多々あることを付け加えておきたい。今回は校正を終了した基本語辞書を研究用に限り公開することにした。

参考文献

- 1) 栗原, 黒崎: 仮名文の漢字混り文への変換について, 九大工学集報, Vol. 39, No. 4, pp. 659-664 (1967).
- 2) 栗原, 稲永: カナ漢字変換[I], 九大工学集報, Vol. 42, No. 6 (1970).
- 3) 稲永, 小西: カナ文字文の機械処理のための辞書について, 電子通信学会技報 AL 76-39 (1976).
- 4) 稲永, 小西, 吉田: かな漢字変換用辞書, 昭和54-55年度文部省科研費試験研究(1)「かな漢字変換を中心とした日本語入力システムの開発」研究報告書(代表者吉田), pp. 80-111 (1981).
- 5) 長尾, 辻井, 山上, 建部: 国語辞書の記憶と日本語文の自動分割, 情報処理, Vol. 19, No. 6 (1978).
- 6) 和智: 単語辞書システムの作成——階層化された二進木構造を持つ自立語辞書, 九大工学研究科修士論文(1980).
- 7) 中野, 野村: 日本語の形態素分析, 情報処理, Vol. 20, No. 10, pp. 857-864 (1979).
- 8) 稲永: 機械辞書の誤り検出のための漢字の音訓表作成について, 昭和56年度電気四学会九州支部連合大会講演論文集, No. 34 (1981).
- 9) 吉田, 栗原, 松尾: 日本語の機械処理——構文分析から意味分析へ——, 電子通信学会論文誌(D), Vol. 55-D, No. 6 (1972).
- 10) 首藤: 文節構造モデルによる日本語の機械処理に関する研究, 福岡大学研究所報, No. 45 (1980).
- 11) 稲永, 小山: 文節構造解析のための接続辞書の構成と付属語の接続検定, 九大工学集報, Vol. 54, No. 4, pp. 399-405 (1981).

(昭和56年10月30日受付)