

UniDic version 1.1.0 ユーザーズマニュアル

伝 康晴・山田 篤・宇津呂武仁

平成 15 年 8 月

UniDic version 1.1.0 Users Manual

Yasuharu Den, Atsushi Yamada, and Takehito Utsuro

Copyright © 2002-2003 Chiba University and Information-technology Promotion Agency, Japan (IPA). All rights reserved.

This edition is for UniDic version 1.1.0.

Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies.

Permission is granted to copy and distribute modified versions of this manual under the above conditions for above verbatim copying, provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

Permission is granted to copy and distribute translations of this manual into another language, under the above conditions for modified versions.

この版は UniDic version 1.1.0 に対応する。

著作権表示とこの許可文がすべての複製に存在する限り、この説明書のまったく同一の複製を作り、配布することを許可する。

同一複製の条件の下で、それによって得られた結果もこの許可文の表示と同一の条件のもとで配布する限り、この説明書の修正版の複製をし、配布することを許可する。

上記の修正版に関する条件の下で、この説明書の翻訳の複製を作り、配布することを許可する。

version 1.1.0 21 August 2003

目次

第 I 部	UniDic 茶釜辞書	2
1	インストール	2
1.1	Unix でのインストール	2
1.2	Windows でのインストール	2
2	各ファイルのフォーマット	3
2.1	品詞定義ファイルのフォーマット	3
2.2	活用型定義ファイルのフォーマット	3
2.3	活用形定義ファイルのフォーマット	3
2.4	語彙定義ファイルのフォーマット	4
2.5	複合語定義ファイルのフォーマット	5
2.6	接続規則ファイルのフォーマット	6
3	chasenrc ファイル	6
第 II 部	UniDic 品詞体系	10
4	概要	10
4.1	基本理念	10
4.2	3 つの階層性	11
5	品詞	15
5.1	名詞	15
5.2	代名詞	17
5.3	形状詞	17
5.4	動詞	17
5.5	形容詞	18
5.6	その他の自立語	18
5.7	助動詞	18
5.8	助詞	19
5.9	接頭辞	19
5.10	接尾辞	20
5.11	その他の品詞	21
6	活用型	21
6.1	動詞	21
6.2	形容詞	27

6.3	助動詞	29
6.4	接尾辞	30
7	活用形	30
7.1	語幹	30
7.2	未然形	31
7.3	意志推量形	31
7.4	連用形	31
7.5	基本形	32
7.6	連体形	32
7.7	仮定形	33
7.8	已然形	33
7.9	命令形	33
8	音韻論的属性	34
8.1	語頭音韻交替型	34
8.2	語末音韻交替型	35
8.3	語頭音韻交替結合型	35
8.4	語末音韻交替結合型	36
8.5	アクセント型	36
8.6	アクセント結合型	38
	参考文献	41
付録 A	UniDic を用いた形態素解析	42
付録 B	他の品詞体系との違い	43
B.1	国語研短単位体系との違い	43
B.2	IPA 体系との違い	43
付録 C	To Do リスト	44

はじめに

形態素解析システム『茶筌』(松本, 2000) は、言語処理のためのフリーソフトウェアとして奈良先端科学技術大学院大学より公開されているシステムです。本辞書は、茶筌（version 2.3.3 以降）用の日本語辞書 UniDic について説明したものです。茶筌用辞書としては、奈良先端科学技術大学院大学より公開されている IPA 品詞体系にもとづく日本語辞書 IPADic がありますが、本辞書は、音声情報処理や言語学・国語学分野での利用により適した品詞体系にもとづく辞書です。

本辞書は、情報処理振興事業協会 (IPA) 「擬人化音声対話エージェント基本ソフトウェアの開発」プロジェクト（代表: 東京大学・嵯峨山茂樹教授）の助成を得て作成されました。本辞書の語彙の大半と統計情報は、開放的融合研究「話し言葉工学」の話し言葉コーパス（CSJ 2002 年度モニター版）(前川ほか, 2000)、日本音響学会の研究用連続音声データベース (小林ほか, 1992) の一部、ATR 音素バランス 503 文から取得しました。アクセント情報は、東京大学・広瀬・峯松研究室の協力を得て作成しました。辞書の整備作業は、アルバイトの田中美由紀さん・古川麻衣さんに手伝っていただきました。品詞体系の設計に際しては、奈良先端科学技術大学院大学・松本裕治・浅原正幸、国立国語研究所・小椋秀樹・山口昌也の各氏をはじめ、UniDic メーリングリストに加盟するさまざまな方々から助言をいただきました。高岡一馬さんをはじめ茶筌開発陣の方々には、UniDic の機能を最大限実現するために、さまざまな改良を茶筌に施していただきました。

本辞書の構築に携わられたすべての方々に対して心より感謝します。

本辞書に関するお問い合わせは以下にお願いします。

〒263-8522

千葉市稲毛区弥生町 1-33

千葉大学文学部

伝 康晴

Tel: (043) 290-3577, Fax: (043) 290-2278

E-mail: den@cogsci.L.chiba-u.ac.jp

第 I 部

UniDic 茶釜辞書

第 I 部では、茶釜用の日本語辞書 UniDic の仕様について説明する。UniDic の品詞体系については、第 II 部を参照のこと。また、UniDic を用いた形態素解析については、付録 A も参照。

1 インストール

1.1 Unix でのインストール

本辞書のインストールには、chasen-2.3.3 以降が必要である。インストール手順は以下のとおりである。

1. `./configure` を実行する。

```
% ./configure
```

フィラーを辞書に含めたくない場合は、`--with-exclude-dic` オプションを用いる。

```
% ./configure --with-exclude-dic=fillers.dic
```

2. `make` を実行する。

```
% make
```

3. `root` 権限で `make install` を実行する。デフォルトでは `/usr/local/lib/chasen/dic/unidic` にインストールされる。

```
# make install
```

4. `/usr/local/etc/chasencrc` を編集する。辞書をはじめてインストールするのであれば、インストーラが自動的に `/usr/local/etc/chasencrc` を作成する。そうでない場合は、パッケージ中に生成される `chasencrc` を参考にして、`/usr/local/etc/chasencrc` を編集する必要がある。

1.2 Windows でのインストール

本辞書のインストールには、Windows 用にコンパイルされた chasen-2.3.3 以降が必要である（WinCha ではインストールできない）。インストール手順は以下のとおりである。

1. 茶釜がインストールされているフォルダ（デフォルトでは `c:\Program Files\chasen`）に辞書パッケージをフォルダごとコピーする。
2. 辞書パッケージ中の `Makefile.bat` を実行する（エクスプローラーでダブルクリックすればよい）。フィラーを辞書に含めたくない場合は、パッケージ中の `fillers.dic` を `.dic` 以外の拡張子に改名する。辞書は、`c:\Program Files\chasen\lib\chasen\dic\unidic` にインストールされる。
3. パッケージ中の `chasencrc` を参考にして、`c:\Program Files\chasen\etc\chasencrc` を編集する。

2 各ファイルのフォーマット

2.1 品詞定義ファイルのフォーマット

`grammar.cha` には、品詞のリストが記述されている。活用のある品詞の場合、品詞名の末尾に % をつける。活用のある品詞では、`ctypes.cha` に可能な活用型を、`cforms.cha` に可能な活用形を記述する必要がある。

```
(助詞
  (格助詞
    (一般)
    (連体可能)
    (連用可能))
  (副助詞)
  (係助詞)
  (接続助詞)
  (終助詞)
  (準体助詞))

(動詞
  (一般 %)
  (非自立可能 %))
```

2.2 活用型定義ファイルのフォーマット

`ctypes.cha` には、活用のある品詞がどのような活用型を取るかが記述されている。

```
((動詞 一般)
  (五段-力行-一般-一般
   五段-力行-一般-イ段
   五段-力行-イク
   五段-力行-ユク
   ...
   文語力行変格-一般
   文語力行変格-く
   文語サ行変格-ス
   文語サ行変格-ズ
   文語ナ行変格
   文語ラ行変格
   無変化型))
```

2.3 活用形定義ファイルのフォーマット

`cforms.cha` には、各活用型がどのような活用形を取り、各活用形がどのような活用語尾を取るかが記述されている。活用語尾は、書字形・仮名書字形・発音形（4.2.2 節参照）の順に記述する。ファイルの冒頭で、基本形の名称を宣言する。

(基本形 "基本形-一般")

(五段-力行-一般-一般

((未然形-一般	か	カ	カ)
((意志推量形-一般	こう	コウ	コー)
((意志推量形-一般:短縮	こ	コ	コ)
((意志推量形-促音便	こっ	コッ	コッ)
((連用形-一般	き	キ	キ)
((連用形-イ音便	い	イ	イ)
((基本形-一般	く	ク	ク)
((仮定形-一般	け	ケ	ケ)
((仮定形-融合	きゃ	キャ	キャ)
((命令形-一般	け	ケ	ケ)))

2.4 語彙定義ファイルのフォーマット

lexicon.dicには、単語のリストが記述されている。ただし、話し言葉に固有であるフィラーについては、fillers.dicに分けて記述してある。書き言葉を対象とする際には、1節に述べた方法でfillers.dicを辞書から除外すればよい。

(品詞 (助詞 副助詞))

((見出し語 (だけ 0)) (読み ダケ) (発音 ダケ)
(付加情報 'lex="ダケ!だけ" indexForm="ダケ" indexOrth="だけ"
aConType="形容詞 %F1, 名詞 %F5, 動詞 %F1"))

(品詞 (動詞 一般))

((見出し語 (書く 769)) (読み カク) (発音 カク)
(活用型 五段-力行-一般-一般)
(付加情報 'lex="カ*ク!書く" indexForm="カク" indexOrth="書く"
aType="1"))

語彙定義には、以下の情報が含まれる (各情報の意味は4.2.2節参照)。

- 品詞: 品詞名中の‘-’を空白で置き換えたものが記述される。
- 見出し語: 書字形と単語生起コストの対が記述される。
- 読み: 仮名書字形が記述される。
- 発音: 発音形が記述される。
- 活用型: 活用のある語の活用型が記述される。活用型は `ctypes.cha` に定義されたもののみに許される。
- 付加情報: 語概念・語形・表記と音韻論的属性がまとめて記述される。各属性の値は、「属性="値"」の形式で記述される (属性値対の列全体が’で囲まれていることに注意)。可能な属性は以下のとおり。
 - lex: 語概念
 - indexForm: 語形
 - indexOrth: 表記
 - ipaType: 語頭音韻交替型
 - ipaConType: 語頭音韻交替結合型
 - fpaType: 語末音韻交替型

- fpaConType: 語末音韻交替結合型
- aType: アクセント型
- aConType: アクセント結合型

単語生起コストは、数値が小さいほどその語が出現しやすいという意味である。UniDic 茶筌辞書の単語生起コストは、形態素解析済みコーパスから取得した統計情報にもとづいて付与されている。

2.5 複合語定義ファイルのフォーマット

complex.dic には、複合語のリストが記述されている。複合語は、何らかの理由により、複数の語からなる列を一塊にして解析したい場合に用いる。あたかも一語であるかのように塊として扱われるため、複合語として登録した解析結果を優先的に得ることができる。解析時には塊として扱われても、chasenrce の「複合語出力」オプションに"SEG"を指定することで、個々の語を別々に出力できるので、単位の斉一性を損なわない。

```
(品詞 (名詞 普通名詞 一般))
((見出し語 (和洋中 4000)) (読み ワヨウチュウ) (発音 ワヨーチュー)
(複合語
  ((品詞 (名詞 普通名詞 一般))
   (見出し語 和) (読み ワ) (発音 ワ)
   (付加情報 'lex="ワ!和" indexForm="ワ" indexOrth="和"
              aType="1" aConType="C3"))))
  ((品詞 (名詞 普通名詞 一般))
   (見出し語 洋) (読み ヨウ) (発音 ヨー)
   (付加情報 'lex="ヨウ!洋" indexForm="ヨウ" indexOrth="洋"
              aType="1" aConType="C3"))))
  ((品詞 (名詞 普通名詞 一般))
   (見出し語 中) (読み チュウ) (発音 チュー)
   (付加情報 'lex="チュウ!中" indexForm="チュウ" indexOrth="中"
              aType="1" aConType="C4")))) ))

(品詞 (名詞 普通名詞 一般))
((見出し語 (着心地 4000)) (読み キゴコチ) (発音 キゴコチ)
(複合語
  ((品詞 (動詞 一般))
   (見出し語 着) (読み キ) (発音 キ)
   (原形 着る) (活用型 上二段-力行) (活用形 連用形-一般)
   (付加情報 'lex="キル!着る" indexForm="キル" indexOrth="着る"
              aType="0"
              w2Chunk="B"
              w2_orth="着心地" w2_pron="キゴコチ"
              w2_pos="名詞-普通名詞-一般"
              w2_aType="0,2" w2_aConType="C2"))))
  ((品詞 (名詞 普通名詞 一般))
   (見出し語 心地) (読み ココチ) (発音 ココチ)
   (付加情報 'lex="ココチ!心地" indexForm="ココチ" indexOrth="心地"
              aType="0" aConType="C2"
              w2Chunk="E')))) ))
```

複合語定義では、通常の語彙定義に続けて、複合語フィールドに構成要素の情報を記述する。各構成要素は、それ自体、語彙定義の形式をとる。ただし、構成要素が活用語の場合は、活用した形で見出し語・読み・発

音を記述し、あわせて原形と活用形を記述する。なお、上の 2 番目の例では、構成要素の付加情報に **w2Chunk** や **w2_orth** といった **w2** ではじまる名前の属性がいくつか記述されているが、これらは第 2 層語 (4.2.1 節参照) に関する情報であり、後段のモジュールによって XML による階層的な表現に整形される。

複合語を使う動機は、以下のようなものが考えられる。

- 別々の語としては正しく解析できない。
- 連濁や複合語アクセントなどが正しく扱えないため、第 2 層語を直接解析したい。

上記の最初の例は前者に、2 番目の例は後者に相当する。

2.6 接続規則ファイルのフォーマット

connect.cha には、接続規則が記述されている。接続規則とは、ある要素とある要素がどれくらいつながりやすいかを規定したものである。

```
(( (((名詞 固有名詞 人名 姓)))
  (((接尾辞 名詞的 人名))) )
806)

(( (((動詞 一般) 五段-力行-一般-一般 未然形-一般))
  (((助動詞) 助動詞-ナイ * ない)) )
344)

(( (((*))
  (((助動詞) 助動詞-デス * です)))
8000)

(( (((助詞 準体助詞) * * ん))
  (((助動詞) 助動詞-デス * です)) )
269)
```

接続規則は、接続情報と接続コストの対からなり、接続情報は、前件要素定義と後件要素定義の対からなる (trigram 規則では、前々件・前件・後件の 3 つ組)。各要素定義は、品詞・活用型・活用形・書字形の 4 つ組であり、'*' をワイルドカードとして用いることができる。4 つ組の後部に '*' が連続する場合は省略できる。

接続コストは、数値が小さいほど前件要素と後件要素がつながりやすいという意味である。UniDic 茶筌辞書の接続コストは、形態素解析済みコーパスから取得した統計情報にもとづいて付与されている。

ファイルの前方にある接続規則は、ファイルの後方にある規則によって上書きされる。このため、一般的な規則を先に書き、詳細化した規則を後に書く必要がある。たとえば、上例の 3 番目の規則は 4 番目の規則によって上書きされるため、「助詞-準体助詞 ん」と「助動詞 です」の接続に際しては、接続コスト 269 が、「助詞-準体助詞 ん」以外と「助動詞 です」の接続に際しては、接続コスト 8000 が用いられる。

3 chasenrc ファイル

chasenrc ファイルは、茶筌の実行に必要なさまざまなオプションが定義されている。以下のような優先順位で、chasenrc ファイルが読み込まれる。

1. 起動時に-r オプションで指定されたファイル
2. 環境変数 CHASENRC で指定されたファイル
3. ユーザのホームディレクトリにある.chasenrc
4. /usr/local/etc/chasenrc

オプションの一覧を以下に示す。

1. 文法ファイルのディレクトリ

grammar.cha, ctypes.cha, cforms.cha, connect.cha が存在するディレクトリを指定する。

(文法ファイル /usr/local/lib/chasen/dic/unidic)

2. システム辞書

インストール時に作成されるシステム辞書 (chadic.{da,lex,dat}) のパスを指定する。‘/’ で始まらないパスを記述すると、文法ファイルディレクトリからの相対パスとみなされる。たとえば、以下の記述では、文法ファイルと同じディレクトリにある chadic.{da,lex,dat} が読み込まれる。

(DADIC chadic)

3. 未知語の品詞

未知語が発見された時に、その語をどのような品詞として扱って接続規則を適用するかを指示する。

(未知語品詞 (名詞 普通名詞 一般))

4. 未知語のコスト

未知語の単語生起コストを定義する。

(品詞コスト
((*) 1)
((UNKNOWN) 30000))

5. 未定義接続のコスト

接続規則ファイルに接続規則が定義されていない単語間の接続コストを指定する。

(未定義接続コスト 10000)

6. 単語生起コストと接続コストの相対的な重み

ある解析候補の総コストは、各単語の単語生起コストと各隣接単語間の接続コストの総和として算出される。これら2種類のコストに異なる重みづけを与えたい場合には、それを指定することができる。

(形態素コスト重み 1)
(接続コスト重み 1)

7. コスト幅

最小コストを持つ解析候補の探索にあたって、探索ステップを進めるたびに、その時点でコストが最小の解だけを保持するのではなく、ある程度のコスト幅を許容して複数解を保持することで、より精密に

最適解を求めることができる。この許容幅を指定することができる（実用上は、コスト幅 0 でもほとんど精度は変わらない）。コスト幅は、起動時の **-w** オプションで上書きできる。

(コスト幅 0)

8. 出力フォーマット

出力フォーマットを指定することにより、解析結果の出力形式を変えることができる。

(出力フォーマット "`<W1 orth=\"%m\" pron=\"%a\"
pos=\"%U(%P-)%T/\" cType=\"%T \"/%?F/%cForm=\"%F \"/
%?I/ %i//>\n`")

上記の指定により、以下のような xml 形式の出力が得られる。

```
<W1 orth="文" pron="ブン"  
pos="名詞-普通名詞-一般"  
lex="ブン!文" indexForm="ブン" indexOrth="文"  
fpaConType="B" aType="1" aConType="C3"/>  
<W1 orth="を" pron="オ"  
pos="助詞-格助詞-一般"  
lex="ヲ!を" indexForm="ヲ" indexOrth="を"  
aConType="動詞 %F5, 名詞 %F1, 形容詞 %F6@0, -2"/>  
<W1 orth="解析" pron="カイセキ"  
pos="名詞-普通名詞-サ変可能"  
lex="カイセキ!解析" indexForm="カイセキ" indexOrth="解析"  
aType="0" aConType="C2"/>  
<W1 orth="する" pron="スル"  
pos="動詞-非自立可能" cType="サ行変格-為ル" cForm="基本形-一般"  
lex="スル!為る" indexForm="スル" indexOrth="為る"  
aType="0"/>
```

フォーマット指定の詳細については、IPADic のマニュアル (浅原・松本, 2003) を参照。

9. BOS 文字列

解析結果の文頭に表示する文字列を指定する。デフォルトは空文字列。

(BOS 文字列 "`<S>\n`")

10. EOS 文字列

解析結果の文末に表示する文字列を指定する。デフォルトは `"EOS\n"`。

(EOS 文字列 "`</S>\n`")

11. 空白品詞

茶筌は、半角のスペースとタブを空白とみなし、これらを見做して解析する。通常は、解析結果に空白の情報を出力しないが、「空白品詞」を設定することにより、空白についての情報を出力するようになる。たとえば、以下のように設定すると、空白の品詞を「その他-空白」として出力する。

(空白品詞 (その他 空白))

12. 注釈

ある文字列で始まり、ある文字列で終わる一つながりの文字列を注釈として扱い、その文字列の部分を解析時に無視させることができる。「注釈」オプションで、注釈の開始文字列・終了文字列と出力フォーマットを指定できる。終了文字列を省略すると、開始文字列と完全一致する文字列を注釈として扱う。また、出力フォーマットを省略すると、注釈についての情報をまったく出力しない。たとえば、以下のように指定すると、“<emotion TYPE="HAPPY">”のように、‘<’と‘>’で囲まれた範囲を解析の対象外とし、該当する位置にそのまま出力する。

(注釈 ("<" ">") "%m\n")

13. 連結品詞

ある品詞の単語が連続して出現したときに、一つの塊として連結して出力させるときに使用する。たとえば、以下の記述では、連続したアラビア数字を一つの塊として出力する。

(連結品詞 ((名詞 数詞 アラビア数字)))

単位の斉一性を損なうので、後処理を前提とするのでない限り、用いないほうがよい。

14. 区切り文字

-j オプション使用時の文の区切り文字を指定する。区切り文字には、全角文字・半角文字の両方を使用することができる。

(区切り文字 "。、。 !?.,!? ")

15. 複合語出力

complex.dic で定義した複合語について、塊を一語として出力するか ("COMPOUND")、構成要素を別々の語として出力するか ("SEG") を指定する。デフォルトは "COMPOUND" であるが、単位の斉一性を損なわないよう、"SEG" を指定するほうがよい。

(複合語出力 "SEG")

第 II 部

UniDic 品詞体系

第 II 部では、UniDic の設計と品詞体系について述べる。UniDic を用いた形態素解析について、付録 A も合わせて参照するとよい。

4 概要

4.1 基本理念

UniDic の基本理念は以下のとおりである (伝ほか, 2002)。

- 言語学・国語学など人文系の言語研究に違和感なく使える。
 - 齊一な単位設定
 - 学校文法に準じた品詞体系の採用
- 音声言語研究に使える。
 - 実情に即した発音情報の記述
 - アクセント型情報の記述
 - 音韻交替・アクセント変化に関わる制約の記述

従来の形態素解析では、解析精度の向上および構文・意味処理での使いやすさといった理由から、個々の単語の認定をかなりアドホックに行なっていた。たとえば、「〇〇的」のような生産的な形容動詞を大量に辞書登録したり、「そうですね」のような慣用的表現を一語として登録したりといったことが、普通に行なわれている。しかし、言語学や国語学では、まずもって、いかにして「語」を認定するかということが重要な課題となる。語の認定が研究者ごとに異なれば、それにもとづいた分析結果の意味も変わってしまうからである。UniDic では、同様の問題意識から体系的で操作的な単位認定規則を策定している、国立国語研究所の短単位 (小椋, 2001) をもとにして、齊一な単位設定を行なった。

同様の問題意識は品詞についてもあてはまる。研究者間で分析結果を広く共有し、相互に比較するためには、なるべく多くの者に受け入れられる標準的な品詞体系を採用する必要がある。このような体系としては、学校文法が候補になる。UniDic では、いずれも学校文法をもとにしている、国語研短単位体系 (小椋, 2001) と IPA 体系 (RWCP, 1998) を参考にして、品詞体系を設計した (これらの体系との違いについては、付録 B を参照)。

さらに、形態素解析は通常、「入力文を部分文字列に分割し、それらに品詞・活用型を割り当てること」と考えられており、分割された部分文字列に対して、正しい語形や発音を与えることは、さほど重視されていない。しかし、テキスト音声合成などの音声言語情報処理においては、各語に正しい発音を与えることが重要課題となる。とくに、促音化や連濁などによる音韻交替の扱いは欠かせない。UniDic では、実情に即した発音情報と、おもに数詞に関わる音韻交替のための情報を記述した。

最後に、音声言語研究においては、アクセント情報の記述も必要になる。電子化辞書でアクセント型を記述しているものは少なく、とくに、公開されている形態素解析用辞書に関しては皆無である。また、ここでも、連濁と同様に、語の複合や文節の構成によるアクセント変化を扱う必要がある。たとえば、「社会」は頭高型、「生活」は平板型であるが、これらが複合して「社会生活」になると、4 モーラ目に核を持つ中高型になる。こ

	神戸	女子	大学	を	訪問	し	た
第 1 層	名詞	名詞	名詞	助詞	名詞	動詞	助動詞
第 2 層	名詞	名詞		助詞	名詞	動詞	助動詞
第 3 層	名詞			助詞	動詞		助動詞

図 1: 多階層単位の例

のようなアクセント変化はかなり規則的であり、それに関わる制約を各語の付加情報として辞書に記述しておけば、形態素解析の後処理としてアクセント変化を扱うことができる。UniDic では、語のアクセント型、および語の複合・文節の構成に伴うアクセント変化に関わる制約を記述している。

4.2 3つの階層性

UniDic では、以下の 3 つの事項に関して、階層性を導入している。

1. 語の単位の階層性
2. 語を定義する属性群の階層性
3. 品詞・活用型・活用形の階層性

4.2.1 語の単位の階層性

4.1 節で、齊一な単位の重要性について述べた。しかし、短単位が唯一の選択肢ではない。短単位は、概ね（漢語の）漢字二字を一語とするような短い単位であり、目的によっては、もっと長い単位が望ましいこともある。たとえば、「言語 | モデル」「言語 | 情報」「言語 | 体系」「言語 | 作品」はいずれも短単位二語からなり、複合した形での使用法の違いを見るうえでは、短単位はふさわしくない。国語研では、このような理由から短単位に加えて、長単位を併用している（前川ほか, 2001）（長単位はより文節に近い）。

UniDic でも、同様の理由により、第 1 層から第 3 層までの 3 段階において語の単位を設定した（伝ほか, 2002）。図 1 に、多階層単位の例をあげる。第 1 層が概ね短単位に、第 3 層が概ね長単位に対応する。第 2 層は、音声言語研究での利用を想定して、独自に設けた。

第 1 層語は、原則として、形態素（意味を持つ最小の単位）が二つ結合したものである。たとえば、/女/ と /子/ という二つの形態素が結合して、| 女子 | という第 1 層語ができる。ただし、第 1 層語の構成規則に関しては、多くの例外規定がある。詳細は、短単位に関する文献（小椋, 2001）を参照のこと。

第 2 層は、ある種の複合語を一語としてとらえる単位である。たとえば、| 女子 | と | 大学 | という二つの第 1 層語が結合して、|| 女子大学 || という第 2 層語ができる。結合規則は概ね以下のように規定される。

1. 「0 個以上の接頭辞 + 1 個以上の名詞・形状詞・動詞・形容詞 + 0 個以上の接尾辞」からなる列で、
2. すべての隣接要素が係り受け関係を持つ。

2 の規定により、第 2 層語は内部構造として左分岐構造しか許さない。たとえば、|| 神戸 | 学院 | 大学 || は第 2 層では一語であるが、|| 神戸 || 女子 | 大学 || は二語である（|| は第 2 層語の境界、| は第 1 層語の境界を示す）。このような規則を採用した理由は、右分岐構造が連濁規則やアクセント変化規則の適用をブロックするという知見があるからである（窪園, 1995）。実際、「神戸学院大学」は一つのアクセント単位にまとまるが（|| コーベ | ガクイン | ダ¹イガク ||）、「神戸女子大学」は、「神戸」「女子大学」という二つのアクセント単位に

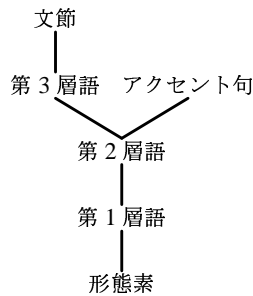


図 2: UniDic における語の単位の階層性

分かれる (|| コ¹ーベ || ジョシ | ダ¹イガク ||)。このような現象を正しく扱うために、連濁規則やアクセント変化規則の適用範囲を規定する単位を設けるのが望ましいと考え、第 2 層を設けた。

第 3 層は、概ね、文節を自立語と付属語に分けたものである。「神戸女子大学」は第 3 層では一語である。また、「サ変名詞」+「する」も一語として扱う。第 3 層語は、統語・意味・語用論的な分析・処理での利用を想定している。

これらの単位と、形態素・アクセント句・文節との関係を図 2 に示す。図からわかるように、第 1 層～第 3 層はそれぞれ、語彙・形態論、音韻論、統語・意味・語用論とのインターフェースである。

UniDic 茶筌辞書は、このうち第 1 層語に関する記述を与えたものである。すなわち、UniDic を用いた茶筌は、与えられた入力文に対して、第 1 層語の列を出力する。第 2 層語・第 3 層語は、その結果を受け、チャンキング処理によって、ある程度自動的に派生できる（ただし、例外的なものについては、茶筌辞書の複合語定義により、直接的に解析する必要がある。2.5 節参照）。

以下、本マニュアルでは、UniDic 第 1 層の体系について説明する。以下でたんに語という場合には、第 1 層語を指すものとする。

4.2.2 語を定義する属性群の階層性

UniDic では、語をさまざまな属性によって定義する。これらの属性の継承関係によって、語を以下のように階層化することができる。

1. 語彙素: 〈語概念〉
2. 索引語: 語彙素を定義する属性群 + 〈語形〉
3. (a) 書字語: 索引語を定義する属性群 + 〈書字形〉
(b) 発音語: 索引語を定義する属性群 + 〈発音形〉

語の定義の階層性の例を図 3 に挙げる。

語彙素 (lemma) は、意味・文法機能を同じくする語の集合に対応するレベルであり、意味と文法機能を対にした“アナタ! 貴方! 代名詞”のような語概念によって定義される。意味は、同じ語彙素に属する語群の中でもっとも代表的なものの語形と表記の対によって表わし（活用のある語では、語幹・語尾の境界を“*”で示す）、文法機能は、その品詞によって表わす。UniDic 茶筌辞書の lex 属性はこのうち意味の部分のみを表わしている。このレベルでは、語形の違いは考慮されず、たとえば、異形態 (allomorph) である「アナタ」と「アanta」はいずれも同じ語概念“アナタ! 貴方! 代名詞”を持つ。

語概念	語形	書字形	発音形
アナタ! 貴方! 代名詞	アナタ	貴方	アナタ
		あなた	アナタ
	アンタ	あんた	アンタ
アラワ*ス! 表わす! 動詞-一般	アラワス	表わす	アラワス
		表す	アラワス
		現わす	アラワス
		現す	アラワス
		あらわす	アラワス
アラワ*ス! 著わす! 動詞-一般	アラワス	著わす	アラワス
		著す	アラワス
		あらわす	アラワス
テ! て! 助詞-接続助詞	テ	て	テ
	デ [†]	で	デ
デ! で! 助詞-接続助詞	デ [‡]	で	デ
カナ! 仮名! 名詞-普通名詞-一般	カナ	仮名	カナ
		かな	カナ
カメイ! 仮名! 名詞-普通名詞-一般	カメイ	仮名	カメー
		かめい	カメー
ウインドウズ! 名詞-固有名詞-一般	ウインドウズ	ウインドウズ	ウインドーズ
			ウィンドーズ
		ウィンドウズ	ウインドーズ
			ウィンドーズ

[†]「飲んできた」の「で」 [‡]「～ならでは」の「で」

図 3: 語の定義の階層性の例

索引語 (lexeme) は、国語辞典の見出しに相当するレベルであり、同じ語彙素に属する語群を語形の違いによって細分化したものである。たとえば、「アナタ」と「アンタ」はこのレベルで別の語として区別される。異なる語彙素に属する語が同じ語形を持つ場合があることに注意しよう。たとえば、「アラワス (表わす)」と「アラワス (著わす)」は同じ語形を持つが、属する語彙素が異なるため、異なる索引語である。同様に、「飲んできた」の「で」と「～ならでは」の「で」は同じ語形を持つが、属する語彙素が異なるため、異なる索引語である (前者は「テ」の異形態)。

書字語 (orthographical) は、テキストに書字されたものとしての語であり、索引語を表記の違いによって細分化したものである。国語辞典の索引に語形とともに併記されている表記 (索引表記) と区別するため、UniDic では、このレベルでの表記を書字形とよぶ。たとえば、「貴方」と「あなた」や「表わす」「表す」「現わす」「現す」「あらわす」などはこのレベルで区別される。語概念の違い (「表わす」と「著わす」と見るか、書字形の違い (「表わす」と「現わす」と見るかは微妙であるが、国語辞典で同一索引の中でも特定の表記に固有の項目が立てられている場合は、その表記を持つものを別の語概念とした (たとえば、小学館『国語大辞典』では、「表わす」「現わす」「著わす」は同じ索引のもとにあるが、「著わす」に固有の項目がある)。異なる索引語に属する語が同じ書字形を持つ場合があることに注意しよう。たとえば、「仮名」という書字形は、「カナ」「カメイ」の二つの索引語に対して可能である。

表 1: UniDic で用いられる属性

階層	定義属性	性質を記述する属性
語彙素	語概念	
索引語	語形	表記, 品詞, 活用型
書字語	書字形	活用型書字形分類, 仮名書字形
発音語	発音形	活用型発音形分類, 語頭音韻交替型, 語頭音韻交替結合型, 語末音韻交替型, 語末音韻交替結合型, アクセント型, アクセント結合型

発音語 (pronunciational) は、音声に発音されたものとしての語であり、索引語を発音の違いによって細分化したものである。たとえば、「ウインドーズ」と「ウィンドーズ」や「データ」と「データー」などはこのレベルで区別される。書字語と発音語は独立に細分化されることに注意しよう。たとえば、「ウインドウズ」と書かれたものを「ウィンドーズ」と読んだり、「ウィンドウズ」と書かれたものを「ウインドーズ」と読んだりすることがあるためである。発音形は、現代日本語の実情に即して、同一母音の連続や [ei], [ou] 連続の 2 モーラ目を長音で表わした（形態素境界をまたぐ場合を除く）（「空白」→「クーハク」、「学校」→「ガッコー」、「先生」→「センセー」）。

以上のような語を定義する属性に加え、UniDic では、各単語の語彙・形態・音韻論的な性質を記述するための属性を設けている。表 1 にそれらをまとめる。

索引語の表記は、国語辞典の索引に併記される漢字と仮名による表現に相当する。表記は、小学館『国語大辞典』などを参考にして定めた。品詞・活用型は次節以降に詳述する。

書字語の活用型書字形分類は、その語の書字形の違いによって、活用語尾の書字形を区別する必要がある場合に用いる。詳細は 6 節で述べる。仮名書字形は、書字形をすべてカタカナで表記したものである。仮名漢字変換システムへの入力をイメージしてもらえばよい。

発音語の活用型発音形分類は、その語の発音形の違いによって、活用語尾の発音形を区別する必要がある場合に用いる。詳細は 6 節で述べる。その他の属性は音韻論的な性質を記述したものである。詳細は 8 節で述べる。

4.2.3 品詞・活用型・活用形の階層性

書字語・発音語が実際に文の中で用いられる際には、活用 (inflection) を伴う場合がある。活用があるかどうかはまず品詞によって区別され、活用がある場合にどのような活用形をとりうるかは活用型によって区別される。品詞・活用型・活用形の詳細はそれぞれ、5, 6, 7 節で述べる。

UniDic の品詞・活用型・活用形は概ね学校文法に準拠している。これは、4.1 節で述べたように、研究者間で分析結果を広く共有し、相互に比較するためには、なるべく多くの者に受け入れられる標準的な品詞体系を採用する必要があるからである。しかし、学校文法の品詞・活用型・活用形は、自然言語処理などの工学的応用の上では分類が粗すぎる。とくに、茶筌自体の高精度な解析のためには、もっと詳細な分類が必要である。

そこで、UniDic では、学校文法の品詞・活用型・活用形を細分化し、さらにそれらを必要に応じて細分化し、といった具合に階層的に品詞・活用型・活用形を設計している。たとえば、「山田」という人名には、「名詞-固有名詞-人名-姓」という 4 レベルまで階層化された品詞が与えられる（‘-’ は階層の区切りを示す）。また、「聞く」という動詞には、「五段-カ行-一般-イ段」という 4 階層の活用型が与えられる。利用者は目的に応じて、必要なレベルまでの分類を参照すればよい。

5 品詞

品詞は一般に、「大分類 [-中分類 [-小分類 [-細分類]]]」の形式をとる。[...] 内は省略される場合がある。各階層の分類の一覧を表 2 に示す。以下、主なものを説明する。

5.1 名詞

■**名詞-普通名詞-サ変可能** 普通名詞のうち、形式的な意味の「いたす」「する」「できる」「なさる」などが後続して、複合的な動詞を作るもの。これらの形式動詞との接続の可能性を示すものであって、実際に後続しているかどうかに関わらずこの品詞を与える。

例: 「アクセス」「悪化」「意図」

■**名詞-普通名詞-形状詞可能** 普通名詞のうち、「な」を伴って連体修飾語になったり、「だ」を伴って述語になったりするもの。可能性を示すものであって、実際に伴っているかどうかに関わらない。

例: 「インターナショナル」「安全」「可能」

■**名詞-普通名詞-副詞可能** 普通名詞のうち、助詞を伴わずに連用修飾語になるもの。句や節による連体修飾を受けて連用修飾節になるものも含む。可能性を示すものであって、実際に助詞を伴わずに連用修飾しているかどうかに関わらない。

例: 「今日」「以上」「時」「為」「通り」

■**名詞-普通名詞-一般** 上記以外の普通名詞。

例: 「A I」「コンピュータ」「花火」

■**名詞-固有名詞-人名-姓** 人名の姓。

例: 「クリントン」「山田」「金」

■**名詞-固有名詞-人名-名** 人名の名。

例: 「マイケル」「真紀子」「正日」

■**名詞-固有名詞-人名-一般** 姓と名に分けられない人名。通称名・芸名・しこ名などを含む。

例: 「キリスト」「ショーケン」「卑弥呼」

■**名詞-固有名詞-地名-国** 国の名前。

例: 「アメリカ」「日本」「韓国」

表 2: 品詞分類

大分類	中分類	小分類	細分類
名詞	普通名詞	一般 サ変可能 形状詞可能 副詞可能	
		固有名詞	
		一般 人名	一般 姓名
		地名	一般 国
		組織名	
	数詞	一般 アラビア数字	
代名詞			
形状詞	一般 助動詞語幹		
連体詞			
副詞			
接続詞			
感動詞	一般 フィラー		
動詞	一般 非自立可能		
形容詞	一般 非自立可能		
助動詞			
助詞	格助詞	一般 連用可能 連体可能	
	副助詞		
	係助詞		
	接続助詞		
	終助詞		
	準体助詞		

大分類	中分類	小分類	細分類
接頭辞	名詞接続	一般 数詞	
	形状詞接続		
	動詞接続		
	形容詞接続		
接尾辞	名詞的	一般 サ変可能 形状詞可能 副詞可能 人名 地名 組織名 助数詞	
	形状詞的		
	動詞的		
	形容詞的		
記号	一般 文字 数字		
その他	一般 空白 句点 読点 括弧開 括弧閉		

■名詞-固有名詞-地名-一般 国名以外の地名

例: 「カリフォルニア」「東京」「新宿」

■名詞-固有名詞-組織名 大学・企業などの組織の名前。ただし、「東京 | 大学」など第 1 層語の範囲を超えるものは認定せず、単独で第 1 層語を構成するものに限る。

例: 「IBM」「ソニー」「東大」

■名詞-固有名詞-一般 上記以外の固有名詞。元号・屋号・ソフトウェア名などを含む。単独で第 1 層語を構成するものに限る。

例: 「JUMAN」「ウインドウズ」「平成」

■名詞-数詞-アラビア数字 アラビア数字。

例: 「0」「1」「2」

■名詞-数詞-一般 アラビア数字以外の数字。「～十」「～百」「～千」や「幾」「何」「数」を含む。

例: 「一」「二十」「何百」「数千」「万」

5.2 代名詞

■代名詞 代名詞。「それ」+「は」が融合した「そりゃ」なども含む。

例: 「私」「それ」「そりゃ」

5.3 形状詞

■形状詞-助動詞語幹 学校文法で助動詞とされる「そうだ（伝聞）」「ようだ」「みたいだ」の語幹部分。

例: 「そう」「よう」「みたい」

■形状詞-一般 学校文法で形容動詞とされるものの語幹部分。ただし、名詞としての用法があるものは、「名詞-普通名詞-形状詞可能」に分類される。

例: 「スマート」「穏やか」「奇妙」

5.4 動詞

■動詞-非自立可能 動詞のうち、「名詞-普通名詞-サ変可能」に接続する「いたす」「する」「できる」「なさる」の類や、動詞連用形やテ形に接続して補助的な意味を添える「始める」「続ける」「くる」「おく」の類。可能性を示すものであって、実際に補助的に使われているかどうかに関わらない。

例: 「する」「始める」「くる」

■動詞-一般 上記以外の動詞。

例: 「聞く」「来たる」「愛する」

5.5 形容詞

■形容詞-非自立可能 形容詞のうち、形容詞・助動詞連用形やテ形に接続して補助的な意味を添える「ない」「ほしい」「よい」の類。可能性を示すものであって、実際に補助的に使われているかどうかに関わらない。

例: 「ない」「ほしい」「よい」

■形容詞-一般 上記以外の形容詞。

例: 「小さい」「ちっちゃい」「少なし」

5.6 その他の自立語

■連体詞 連体詞。

例: 「あの」「大きな」「同じ」

■副詞 副詞。擬態語を含む。ただし、名詞としての用法があるものは、「名詞-普通名詞-副詞可能」に分類される。

例: 「しっかり」「決して」「ピョンピョン」

■接続詞 接続詞。

例: 「しかし」「じゃ」「かつ」

■感動詞-フィラー 話し言葉に固有のフィラー。

例: 「あの」「えーと」「えーっと」

■感動詞-一般 フィラー以外の感動詞。

例: 「ええ」「はい」「ありゃ」

5.7 助動詞

■助動詞 助動詞。接続助詞「て」+「おく」が融合した「とく」なども含む。なお、UniDicでは、学校文法でいう意志・推量の助動詞「う」「よう」を活用形（意志推量形）として扱うので、これらの語はない。

例: 「です」「とく」「べし」

5.8 助詞

■助詞-格助詞-連用可能 格助詞のうち、副詞に接続して「しっかり|と」「余り|に」などの連用修飾句を作るもの。可能性を示すものであって、実際に副詞に接続しているかどうかに関わらない。

例: 「と」「っと」「に」

■助詞-格助詞-連体可能 格助詞のうち、名詞・代名詞に接続して「私|の」などの連体修飾句を作るもの。可能性を示すものであって、実際に連体修飾しているかどうかに関わらず、連用修飾（「私の書いた本」）している場合でもこの分類を用いる。

例: 「の」「ん」

■助詞-格助詞-一般 上記以外の格助詞。「で」+「は」が融合した「じゃ」なども含む。

例: 「が」「で」「じゃ」

■助詞-副助詞 副助詞。「しか」が縮約した「っきゃ」なども含む。

例: 「って」「しか」「っきゃ」

■助詞-係助詞 係助詞。

例: 「は」「も」「こそ」「ぞ」「や」

■助詞-接続助詞 接続助詞。

例: 「て」「けれど」「けど」

■助詞-終助詞 終助詞。

例: 「よ」「ね」「ねん」

■助詞-準体助詞 準体助詞。いわゆる形式名詞の「の」。

例: 「の」「ん」

5.9 接頭辞

■接頭辞-名詞接続-数詞 数詞に接続する接頭辞。

例: 「第（一）」「約（三名）」「計（五回）」

■接頭辞-名詞接続-一般 数詞以外の名詞や動詞連用形に接続する接頭辞。

例: 「御（名前）」「各（家庭）」「前（首相）」「お（食べになる）」

■接頭辞-形状詞接続 形状詞に接続する接頭辞。

例: 「しち(面倒)」

■接頭辞-動詞接続 動詞に接続する接頭辞。

例: 「相(入れる)」「物(足りる)」

■接頭辞-形容詞接続 形容詞に接続する接頭辞。

例: 「超(面白い)」「くそ(寒い)」

5.10 接尾辞

■接尾辞-名詞的-{サ変可能, 形状詞可能, 副詞可能} 名詞に接続して複合名詞を作る接尾辞のうち、作られた複合名詞がそれぞれ、サ変可能・形状詞可能・副詞可能であるもの。

例: 「(一般)化」「(総合)的」「(皮)ごと」

■接尾辞-名詞的-{人名, 地名, 組織名} 名詞に接続して複合的な固有名詞を作る接尾辞のうち、作られた複合名詞がそれぞれ、人名・地名・組織名であるもの。

例: 「(山田)様」「(富士)山」「(東京)大」

■接尾辞-名詞的-助数詞 助数詞。単位を表わす「グラム」「メートル」なども含む。

例: 「(一)個」「(三)つ」「(五)メートル」

■接尾辞-名詞的-一般 上記以外の複合名詞を作る接尾辞。

例: 「(研究)員」「(言語)学」「(三つ)組」

■接尾辞-形状詞的 名詞や動詞連用形に接続して複合的な形状詞を作る接尾辞。学校文法で様態の助動詞とされる「そうだ」の語幹を含む。

例: 「(友達)思い」「(遅れ)がち」「(負け)そう」

■接尾辞-動詞的 動詞連用形や形容詞語幹に接続して複合的な動詞を作る接尾辞。

例: 「(食べ)すぎる」「(でき)かねる」「(痛)がる」

■接尾辞-形容詞的 名詞や動詞連用形に接続して複合的な形容詞を作る接尾辞。

例: 「(信じ)がたい」「(男)らしい」「(外人)っぽい」

5.11 その他の品詞

■記号-文字 アルファベットやギリシャ文字。

例: 「A」「α」「Σ」

■記号-数字 小数点。

例: 「.」

■記号-一般 上記以外の記号。

例: 「ダッシュ」「ハット」「プラス」

■その他-(空白, 句点, 読点, 括弧開, 括弧閉, 一般) 補助記号の類

例: 「」「。」「\」「(」「)」「・」

6 活用型

6.1 動詞

活用型は一般に、「大分類 [-行分類 [-段分類]][-小分類][-書字形分類][-発音形分類]」の形式をとる。[...]内は省略される場合がある。各階層の分類の一覧を表 3-6 に示す。以下、注意すべきものを説明する。

6.1.1 五段活用

■五段-カ行-イク カ行五段活用動詞のうち、「イク」の語形をとるもの。連用形の音便形として促音便をとる。

例: 「行く」「逝く」

■五段-カ行-ユク カ行五段活用動詞のうち、「ユク」の語形をとるもの。連用形の音便形を持たない。

例: 「行く」

■五段-カ行-一般 上記以外のカ行五段活用動詞。連用形の音便形としてイ音便をとる。語幹がイ段で終わる語は、連用形イ音便の発音形で長音化が生じる（「キク」→「キー（タ）」）。よって、語幹末がイ段かどうかを発音形分類で分類する。

「五段-カ行-一般-イ段」の例: 「聞く」「引く」「導く」

■五段-ガ行 ガ行五段活用動詞のうち、語幹がイ段で終わる語は、連用形イ音便の発音形で長音化が生じる（「ヒシグ」→「ヒシー（ダ）」）。よって、語幹末がイ段かどうかを発音形分類で分類する。

「五段-ガ行-一般-イ段」の例: 「拉ぐ」

表 3: 活用型分類（口語動詞）

大分類	行分類	段分類	小分類
五段	カ行		一般 イク ユク
		ガ行	
		サ行	
		タ行	
		ナ行	
		バ行	
		マ行	一般 済ム
	ラ行		一般 アル サル
		ワア行	一般 アウ カウ ガウ タウ ダウ ナウ ハウ バウ マウ ヤウ ラウ ワウ
		イ段	一般 イウ
		ウ段	一般 ツウ
		エ段	
		オ段	

大分類	行分類	段分類	小分類
上一段	ア行		
	カ行		
	ガ行		
	ザ行		
	タ行		
	ナ行		
	ハ行		
	バ行		
	マ行		
	ラ行		一般 リル
下一段	ア行		
	カ行		
	ガ行		
	サ行		一般 セル
	ザ行		
	タ行		
	ダ行		
	ナ行		
	ハ行		
	バ行		
カ行変格	マ行		
	ラ行		一般 レル 呉レル
サ行変格			為ル スル ズル
無変化型			

■五段-マ行-済ム マ行五段活用動詞のうち、「済む」の連用形は「ません」に続く場合に限り、イ音便をとることがある。

■五段-ラ行-サル ラ行五段活用動詞のうち、「なさる」「下さる」は、連用形が「ます」に続く場合にイ音便をとり、命令形が「～い」の形をとる。さらに、連用形促音便で、「さ」が「す」に変わるという現象が見られる（「なさっ（た）」→「なすっ（た）」）。

■五段-ラ行-アル ラ行五段活用動詞のうち、「いらっしゃる」「おっしゃる」「ござる」は、「なさる」「下さる」と同様、連用形が「ます」に続く場合にイ音便をとり、命令形が「～い」の形をとる。

■**五段-ワア行-ア段-〇ウ** ワア行五段活用動詞のうち、語幹がア段で終わる語は、連用形ウ音便で語幹末がア段からオ段に変わる（「ア・ウ」→「オ・ウ（タ）」）。語幹が漢字で書かれる場合には、この変化は書字形には及ばないが（「会う」→「会う（た）」）、語幹が仮名で書かれる場合には、この変化は書字形にも及ぶ（「あう」→「おう（た）」）。よって、語幹末が仮名で書かれるかどうかを書字形分類で分類する。

「五段-ワア行-ア段-アウ-あう」の例：「あう」「取りあう」「触れあう」

■**五段-ワア行-イ段-イウ** ワア行五段活用動詞のうち、「イウ」の語形を持つもの。連用形の発音形で長音化（「イー（マス）」）が、基本形で拗音化（「ユウ」）が見られる。

例： 「言う」

■**五段-ワア行-ウ段-ツウ** ワア行五段活用動詞のうち、「(ッ) ツウ」の語形を持つもの。基本形に「てえ」の形をとる。

例： 「つう」「つつう」

6.1.2 一段活用

■**上一段-ア行** ア行上一段活用動詞のうち、語幹がイ段で終わる語は、連用形の発音形で長音化が生じる（「モチイル」→「モチー（タ）」）。よって、語幹末がイ段かどうかを発音形分類で分類する。

「上一段-ア行-イ段」の例：「用いる」「強いる」

■**上一段-ラ行-リル** ラ行上一段活用動詞の一部は、未然形に撥音便を持つ（「足ん（ない）」）。

例： 「足りる」「下りる」「飛び降りる」

■**下一段-ア行** ア行下一段活用動詞のうち、語幹がエ段で終わる語は、連用形の発音形で長音化が生じる（「ウレエル」→「ウレー（タ）」）。よって、語幹末がエ段かどうかを発音形分類で分類する。

例： 「憂える」

■**下一段-サ行-セル** サ行下一段活用動詞のうち、「着せる」「見せる」などの使役動詞では、連用形で「せ」が「し」に変わる現象が見られる（「見せ（て）」→「見し（て）」）。

例： 「着せる」「見せる」「乗せる」

■**下一段-ラ行-呉レル** ラ行下一段活用動詞のうち、「呉れる」の命令形は「～よ」や「～ろ」の形ではなく、「呉れ」である。

■**下一段-ラ行-レル** ラ行下一段活用動詞の一部は、未然形に撥音便を持つ（「知ん（ない）」）。

例： 「知れる」「切れる」「現われる」

6.1.3 変格活用

■**カ行変格** カ行変格活用動詞は、漢字で書かれる場合には、活用変化が書字形には及ばないが（「来る」→「来（ない）」）、仮名で書かれる場合には、活用変化が書字形にも及ぶ（「くる」→「こ（ない）」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「カ行変格-くる」の例：「くる」「いでくる」

■**サ行変格-為ル** サ行変格活用動詞のうち、「する」にのみ、未然形で、助動詞「ず」に続く形（「セ（ズ）」）と助動詞「せる」に続く形（「サ（セル）」）の区別がある。

■**サ行変格-スル** サ行変格活用動詞のうち、語形が「スル」で終わるもの。

例：「愛する」「関する」「称する」

■**サ行変格-ズル** サ行変格活用動詞のうち、語形が「ズル」で終わるもの。

例：「論ずる」「生ずる」「感ずる」

6.1.4 文語四段活用

■**文語四段-ハ行-ア段-給ウ** 文語ハ行四段活用動詞のうち、「給う」の基本形にはウ音便がある。

■**文語四段-ハ行-ア段-〇ウ** 文語ハ行四段活用動詞のうち、語幹がア段で終わる語は、連用形ウ音便で語幹末がア段からオ段に変わる（「カ・ウ」→「コ・ウ（タ）」）。語幹が漢字で書かれる場合には、この変化は書字形には及ばないが（「買ふ」→「買う（たり）」）、語幹が仮名で書かれる場合には、この変化は書字形にも及ぶ（「かふ」→「こう（たり）」）。よって、語幹末が仮名で書かれるかどうかを書字形分類で分類する。

「文語四段-ハ行-ア段-カウ-かう」の例：「かふ」

■**文語四段-ハ行-イ段-イウ** 文語ハ行四段活用動詞のうち、「イウ」の語形を持つもの。連用形の発音形で長音化（「イー（マス）」）が、基本形で拗音化（「ユウ」）が見られる。

例：「言ふ」

6.1.5 文語二段活用

■**文語下二段-ア行** 文語ア行下二段活用動詞は、「得」「心得」の二語であるが、これらは「得」の部分の音が活用によって変化する（「ウ」→「エ（ズ）」）。「得」が漢字で書かれる場合には、この変化は書字形には及ばないが（「得」→「得（ず）」）、仮名で書かれる場合には、この変化は書字形にも及ぶ（「う」→「え（ず）」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

■**文語下二段-ダ行** 文語ダ行下二段活用動詞の「出」で、「得」と同様の状況が生じる。

■**文語下二段-ナ行** 文語ナ行下二段活用動詞の「寝」で、「得」と同様の状況が生じる。

表 4: 活用型分類（文語動詞）

大分類	行分類	段分類	小分類
文語四段	カ行		
	ガ行		
	サ行		
	タ行		
	ハ行	ア段	カウ ガウ ナウ 給ウ
		イ段	一般 イウ
		ウ段	
		エ段	
		オ段	
	バ行		
	マ行		
	ラ行		
文語上二段	カ行		
	ガ行		
	タ行		
	ダ行		
	ハ行		
	バ行		
	マ行		
	ヤ行		
	ラ行		
大分類	行分類	段分類	小分類
文語下二段	ア行		
	カ行		
	ガ行		
	サ行		
	ザ行		
	タ行		
	ダ行		
	ナ行		
	ハ行		一般 経
	バ行		
	マ行		
	ヤ行		
	ラ行		
	ワ行		
文語カ行変格			
文語サ行変格			ス ズ
文語ナ行変格			
文語ラ行変格			

■**文語下二段-ハ行-経** 文語ハ行下二段活用動詞のうち、「フ」の語形を持つもの。文語ハ行下二段活用動詞は、通常、語尾に書字でハ行の文字を用いても（「考ふ」）、音としてはア行音である（「カンガウ」）。しかし、「経」だけは、音としてもハ行音である。この語には、「得」と同様の状況が生じる。

6.1.6 文語変格活用

■**文語カ行変格** 文語カ行変格活用動詞は、漢字で書かれる場合には、活用変化が書字形には及ばないが（「来」→「来（ず）」）、仮名で書かれる場合には、活用変化が書字形にも及ぶ（「く」→「こ（ず）」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「文語カ行変格-く」の例：「く」

■**文語サ行変格-ス** 文語サ行変格活用動詞のうち、「す」、および語形が「ス」で終わるもの。文語では、「す」と「～す」に活用の違いがない。

例： 「す」「接す」

表 5: 活用型書字形分類（動詞）

活用型	書字形分類
五段-ワア行-ア段-アウ	一般 あう
五段-ワア行-ア段-カウ	一般 かう
五段-ワア行-ア段-ガウ	一般 がう
五段-ワア行-ア段-タウ	一般 たう
五段-ワア行-ア段-ダウ	一般 だう
五段-ワア行-ア段-ナウ	一般 なう
五段-ワア行-ア段-ハウ	一般 はう
五段-ワア行-ア段-バウ	一般 ばう
五段-ワア行-ア段-マウ	一般 まう
五段-ワア行-ア段-ヤウ	一般 やう
五段-ワア行-ア段-ラウ	一般 らう
五段-ワア行-ア段-ワウ	一般 わう

活用型	書字形分類
カ行変格	一般 くる
文語四段-ハ行-ア段-カウ	一般 かふ
文語四段-ハ行-ア段-ガウ	一般 がふ
文語四段-ハ行-ア段-ナウ	一般 なふ
文語四段-ハ行-ア段-給ウ	一般 まふ
文語下二段-ア行	一般 う
文語下二段-ダ行	一般 づ
文語下二段-ナ行	一般 ぬ
文語下二段-ハ行-経	一般 ふ
文語カ行変格	一般 く

表 6: 活用型発音形分類（動詞）

活用型	発音形分類
五段-カ行-一般	一般 イ段
五段-ガ行	一般 イ段
上一段-ア行	一般 イ段
下一段-ア行	一般 エ段

表 7: 活用型分類（形容詞）

大分類	段分類	小分類
形容詞	ア段	カイ ガイ サイ タイ チャイ ナイ バイ パイ マイ ヤイ ライ ワイ 無イ

大分類	段分類	小分類
形容詞	イ段	一般 良イ
	ウ段	ウイ クイ グイ スイ ズイ ツイ ブイ ムイ ユイ ルイ
	エ段	

大分類	段分類	小分類
形容詞	オ段	オイ コイ ゴイ ショイ ソイ ゾイ トイ ドイ ボイ ポイ モイ ヨイ ロイ 良イ
文語形容詞		ク シク

■**文語サ行変格-ズ** 文語サ行変格活用動詞のうち、語形が「ズ」で終わるもの。

例: 「論ず」

6.1.7 その他

■**無変化型** 口語の動詞「うる（得る）」は基本形しか持たない。

6.2 形容詞

活用型は一般に、「大分類 [-段分類][-小分類][-書字形分類][-発音形分類]」の形式をとる。[...] 内は省略される場合がある。各階層の分類の一覧を表 7-9 に示す。以下、主なものを説明する。

■**形容詞-ア段-無イ** ア段形容詞のうち、「無い」、および「無い」を末尾に持つ複合的な形容詞は、接尾辞「そう」（学校文法でいう様態の助動詞「そうだ」）に接続する際に、「語幹」+「さ」の形をとる（「ナサ（ソウダ）」）。さらに、連用形のウ音便において語幹末がオ段に変化したり（「ナ・イ」→「ノ・ウ」）、基本形のエ形において語幹末がエ段に変化したりし（「ナ・イ」→「ネ・エ」）、漢字で書かれる場合には、これらの変化は書字形に及ばないが（「無い」→「無う」）、仮名で書かれる場合には、書字形に及ぶ（「ない」→「のう」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「形容詞-ア段-無イ-ない」の例: 「ない」「頼りない」「だらしない」

■**形容詞-ア段-オイ** ア段形容詞には、連用形のウ音便において語幹末がオ段に変化する（「タカ・イ」→「タコ・ウ」）、基本形のエ形において語幹末がエ段に変化する（「タカ・イ」→「タケ・エ」）といった特徴がある。語幹が漢字で書かれる場合には、これらの変化は書字形に及ばないが（「高い」→「高う」「高え」）、仮名で書

表 8: 活用型書字形分類（形容詞）

活用型	書字形分類	活用型	書字形分類	活用型	書字形分類
形容詞-ア段-カイ	一般 かい	形容詞-ウ段-ウイ	一般 うい	形容詞-オ段-オイ	一般 おい
形容詞-ア段-ガイ	一般 がい	形容詞-ウ段-クイ	一般 くい	形容詞-オ段-コイ	一般 こい
形容詞-ア段-サイ	一般 さい	形容詞-ウ段-グイ	一般 ぐい	形容詞-オ段-ゴイ	一般 ごい
形容詞-ア段-タイ	一般 たい	形容詞-ウ段-スイ	一般 すい	形容詞-オ段-ショイ	一般 しょい
形容詞-ア段-チャイ	一般 ちゃい	形容詞-ウ段-ズイ	一般 ずい	形容詞-オ段-ソイ	一般 そい
形容詞-ア段-ナイ	一般 ない	形容詞-ウ段-ツイ	一般 つい	形容詞-オ段-ゾイ	一般 ぞい
形容詞-ア段-バイ	一般 ばい	形容詞-ウ段-ブイ	一般 ぶい	形容詞-オ段-トイ	一般 とい
形容詞-ア段-パイ	一般 ぱい	形容詞-ウ段-ムイ	一般 むい	形容詞-オ段-ドイ	一般 どい
形容詞-ア段-マイ	一般 まい	形容詞-ウ段-ユイ	一般 ゆい	形容詞-オ段-ボイ	一般 ぼい
形容詞-ア段-ヤイ	一般 やい	形容詞-ウ段-ルイ	一般 るい	形容詞-オ段-ポイ	一般 ぽい
形容詞-ア段-ライ	一般 らい			形容詞-オ段-モイ	一般 もい
形容詞-ア段-ワイ	一般 わい			形容詞-オ段-ヨイ	一般 よい
形容詞-ア段-無イ	一般 ない			形容詞-オ段-ロイ	一般 ろい
				形容詞-オ段-良イ	一般 よい

表 9: 活用型発音形分類（形容詞）

活用型	発音形分類
形容詞-オ段-オイ	一般 オ段

かれる場合には、書字形に及ぶ（「たかい」→「たこう」「たけえ」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「形容詞-ア段-カイ-かい」の例：「たかい」「あたたかい」「あったかい」

■形容詞-イ段-良イ イ段形容詞のうち、「いい（良い）」は基本形しか用いられない（ただし、「まあいっか」に見られるような促音便がある）。

■**形容詞-イ段-一般** イ段形容詞は、連用形のウ音便において、語幹末が拗音化する（「ウレシ・イ」→「ウレシユ・ウ」）。

■**形容詞-ウ段-オイ** ウ段形容詞は、基本形のエ形で語幹末がエ段に変化する（「サム・イ」→「サメ・エ」）。語幹が漢字で書かれる場合には、この変化は書字形に及ばないが（「寒い」→「寒え」）、仮名で書かれる場合には、書字形に及ぶ（「さむい」→「さめえ」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「形容詞-ウ段-ムイ-むい」の例：「さむい」「ねむい」

■**形容詞-エ段** エ段形容詞は理論的にはありうるが、現在、事例はない。

■**形容詞-オ段-良イ** ア段形容詞のうち、「良い」は、接尾辞「そう」（学校文法でいう様態の助動詞「そうだ」）に接続する際に、「語幹」＋「さ」の形をとる（「ヨサ（ソウダ）」）。さらに、基本形のエ形で語幹末がエ段に変化し（「ヨ・イ」→「エ・エ」）、漢字で書かれる場合には、この変化は書字形に及ばないが（「良い」→「良え」）、仮名で書かれる場合には、書字形に及ぶ（「よい」→「ええ」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「形容詞-オ段-良イ-よい」の例：「よい」

■**形容詞-オ段-オイ** オ段形容詞は、基本形のエ形で語幹末がエ段に変化する（「オモ・イ」→「オメ・エ」）。語幹が漢字で書かれる場合には、この変化は書字形に及ばないが（「重い」→「重え」）、仮名で書かれる場合には、書字形に及ぶ（「おもい」→「おめえ」）。よって、仮名で書かれるかどうかを書字形分類で分類する。

「形容詞-オ段-モイ-もい」の例：「おもい」

■**形容詞-オ段-オイ** 末尾が「オイ」で終わるオ段形容詞では、上記に加え、発音上の分類がある。語幹の最後から二つ目の音がオ段である場合には、語幹末のオの発音は長音となる（「オーイ」）。また、このケースでは、連用形ウ音便の発音形が他のケースと異なる（「青い」→「アオー」 vs. 「多い」→「オーウ」）。よって、このケースを発音形分類で分類する。

「形容詞-オ段-オイ-一般-オ段」の例：「多い」「遠い」

■**文語形容詞-ク** 文語ク活用形容詞。

例： 「寒し」「危うし」「少なし」

■**文語形容詞-シク** 文語シク活用形容詞。

例： 「怪し」

6.3 助動詞

活用型は一般に、「大分類-小分類」の形式をとる。各階層の分類の一覧を表 10 に示す。助動詞の活用は、一般に個別的であるため、語形ごとに分類している。ただし、接続助詞「て」＋「おく」が融合した「とく」など、動詞・形容詞由来の融合形助動詞については、動詞・形容詞型の活用型をとる。

表 10: 活用型分類（助動詞）

大分類	小分類
助動詞	ジャ タ タイ ダ デス ナイ ヌ マス ヤ ヤス ラシイ

大分類	小分類
文語助動詞	キ ケム ケリ ゴトシ ザマス ズ タリ ツ ナリ

大分類	小分類
文語助動詞	ヌ ベシ マジ ム ラシ ラム リ ンス
無変化型	

表 11: 活用形分類

大分類	小分類
語幹	一般 サ
未然形	一般 一般:へ 撥音便 ズ接続 セル接続 補助
意志推量形	一般 一般:短縮 促音便

大分類	小分類
連用形	一般 一般:シ イ音便 ウ音便 促音便 促音便:スッ 撥音便 融合 融合:チャ 補助 ト ニ

大分類	小分類
基本形	一般 一般:チャ ウ音便 促音便 撥音便 エ
連体形	一般 一般:短縮 省略 補助 エ:短縮

大分類	小分類
仮定形	一般 融合 融合:キャ
已然形	一般 補助
命令形	一般 イ コ シ ロ

6.4 接尾辞

「接尾辞-動詞的」「接尾辞-形容詞的」の語はそれぞれ、動詞型・形容詞型の活用型をとる。

7 活用形

活用形は一般に、「大分類-小分類」の形式をとる。各階層の分類の一覧を表 11 に示す。以下、注意すべきものを説明する。

7.1 語幹

■語幹-サ 形容詞「無い」と「良い」（およびこれらを末尾に持つ複合的な形容詞）は、接尾辞「そう」（学校文法でいう様態の助動詞）に接続する際に、「語幹」+「さ」の形（「無さ（そう）」「良さ（そう）」）をとる。

7.2 未然形

■未然形-撥音便 ラ行動詞の一部に、未然形語尾が撥音便になるものがある。

例: 「分かん (ない)」「足ん (ない)」「知ん (ない)」

■未然形-一般:へ 関西方言などでは、助動詞「ます」の未然形に、「せ」が「へ」に変化した形（「まへ (ん)」）がある。

■未然形-{ズ接続, セル接続} サ行変格活用動詞「する」の未然形では、助動詞「ず」に続く形（「せ (ず)」）と助動詞「せる」に続く形（「さ (せる)」）の区別がある。

■未然形-補助 文語形容詞などの未然形に、補助活用がある。

例: 「寒から (ず)」

7.3 意志推量形

UniDic では、学校文法でいう意志・推量の助動詞「う」「よう」を立てず、動詞・形容詞・助動詞の活用的一种として意志推量形を設ける。

例: 「行こう」「食べよう」

■意志推量形-促音便 意志推量形の末尾の「う」が促音化することがある。

例: 「行こっ (か)」「食べよっ (か)」

■意志推量形-一般:短縮 意志推量形の末尾の「う」が縮まることがある。

例: 「行こ」「食べよ」

7.4 連用形

■連用形-イ音便 ラ行五段活用動詞の一部に、連用形イ音便がある。

例: 「なさい (ます)」

■連用形-ウ音便 ワア行五段活用動詞や形容詞に、連用形ウ音便がある。

例: 「(おう (た))」

■連用形-促音便:スッ ラ行五段活用動詞の一部に、連用形促音便で「さ」が「す」に変化したものがある。

例: 「なすっ (た)」

■連用形-一般:シ サ行下一段活用動詞の一部に、連用形で「せ」が「し」に変化したものがある。

例: 「着し(て)」

■連用形-融合 助動詞「だ」の連用形で、後続する係助詞「は」と融合した「じゃ」の形がある。

■基本形-融合:チャ 助動詞「だ」の連用形と係助詞「は」の融合形には、「そんなこっちゃだめだ」に見られるような「ちゃ」の形がある。

■連用形-ト 文語助動詞「たり」の連用形に「と」の形がある。

■連用形-ニ 助動詞「だ」の連用形に「に」の形がある。

■連用形-補助 文語形容詞などの連用形に、補助活用がある。

例: 「若かり(し)」

7.5 基本形

UniDic では、学校文法でいう終止形と連体形を区別せず、両者をまとめて基本形とする。この区別にはしばしば離れた係り受け関係を判定する必要があり、形態素解析の範囲を超えるからである。ただし、終止形と連体形で語形が異なる場合は、基本形とは別に連体形を設ける。

■基本形-ウ音便 文語ハ行四段活用動詞「給う」に、基本形ウ音便（「たもう」）がある。

■基本形-促音便 形容詞の基本形で、末尾の「い」が促音化することがある。

例: 「高っ」「安っ」

■基本形-撥音便 動詞基本形が準体助詞や終助詞の「の」に接続する際に、撥音便化することがある。

例: 「食べん(ので)」

■基本形-一般:チャ 助動詞「だ」の基本形には、「何のこっちゃ」に見られるような「ちゃ」の形がある。

■基本形-エ 形容詞や助動詞「たい」「ない」の基本形に、「～え」で終わる形がある。

例: 「高え」「(～し) てえ」「(～し) ねえ」

7.6 連体形

■連体形-一般 助動詞「だ」や文語動詞・形容詞・助動詞の連体形は、終止形とは異なる語形をとるため、連体形を別に立てる。

例: 「な」「忘るる」「若き」「ごとき」

■**連体形-省略** 動詞連体形が準体助詞「ん」に接続する際に、語尾が省略されることがある。

例: 「食べ（んです）」

■**連体形-一般:短縮** 副助詞「(っ) て」+「言う」が融合した「(っ) っう」で、連体形の末尾の「う」が短縮したものがある。

例: 「(何) っ（んだ）」

■**連体形-エ:短縮** 副助詞「(っ) て」+「言う」が融合した「(っ) っう」で、「～え」で終わる形の連体形の末尾の「え」が短縮したものがある。

例: 「(何) て（んだ）」

■**連体形-補助** 文語形容詞などの連体形に、補助活用がある。

例: 「若かる」

7.7 仮定形

■**仮定形-融合** 形容詞や助動詞「たい」「ない」「らしい」の仮定形が接続助詞「ば」と接続するときに、「けりゃ」の形で融合することがある。

例: 「楽しけりゃ」「(～し) たけりゃ」「(～し) なけりゃ」

■**仮定形-融合:キャ** 形容詞や助動詞「たい」「ない」「らしい」の仮定形が接続助詞「ば」と接続するときに、「きゃ」の形で融合することがある。

例: 「楽しきゃ」「(～し) たきゃ」「(～し) なきゃ」

7.8 已然形

■**已然形-補助** 文語形容詞の已然形には、補助活用がある。

例: 「若かれ（ば）」

7.9 命令形

■**命令形-イ** ラ行五段活用動詞の一部の命令形に、「い」で終わる形がある。

例: 「なさい」「下さい」「いらっしゃい」

■**命令形-ロ** 一段活用・サ行変格活用動詞の命令形のうち、「ろ」で終わる形。ラ行五段活用動詞の一部にも「ろ」で終わる形がある。

例: 「食べろ」「しろ」「しゃべろ」

表 12: 語頭音韻交替型分類

語頭音韻 交替型	可能な 語頭の形	この型を とる語の例	語頭音韻 交替型	可能な 語頭の形	この型を とる語の例
カ濁	基本形 (カ) 濁音形 (ガ)	階 (カイ)	へ半濁	基本形 (へ) 半濁音形 (ぺ) 清/半濁音形 (へ/ぺ)	編 (ヘン)
ケ濁	基本形 (ケ) 濁音形 (ゲ)	軒 (ケン)	へ混合	基本形 (へ) 半濁音形 (ぺ) 清/半濁音形 (へ/ぺ) 半濁/濁音形 (ぺ/べ)	遍 (ヘン)
ソ濁	基本形 (ソ) 濁音形 (ゾ)	足 (ソク)	ホ半濁	基本形 (ホ) 半濁音形 (ボ) 清/半濁音形 (ホ/ボ)	歩 (ホ) 報 (ホウ)
ハ半濁	基本形 (ハ) 半濁音形 (パ)	拍, 泊 (ハク) 発 (ハツ)	ホ混合	基本形 (ホ) 半濁音形 (ボ) 濁音形 (ボ)	本 (ホン)
ハ混合	基本形 (ハ) 半濁音形 (パ) 濁音形 (バ)	杯 (ハイ)	ワ混合	基本形 (ワ) 半濁音形 (バ) 濁音形 (バ) 清/半濁音形 (ワ/バ) 清/濁音形 (ワ/バ)	羽 (ワ)
ヒ半濁	基本形 (ヒ) 半濁音形 (ピ)	品 (ヒン)			
ヒ混合	基本形 (ヒ) 半濁音形 (ピ) 濁音形 (ビ)	匹 (ヒキ)			
フ半濁	基本形 (フ) 半濁音形 (プ)	分 (フン)			

■命令形-コ 文語力行変格活用動詞の命令形に、「こ」の形がある。

■命令形-シ 助動詞「ます」の命令形に、「まし」の形がある。

■命令形-一般 上記以外の命令形。

例: 「食べよ」「来い」「せよ」

8 音韻論的属性

UniDic では、音韻論的属性として、数詞がらみの音韻交替やアクセントに関わる属性を記述している。以下、これらを説明する。

8.1 語頭音韻交替型

二つの第 1 層語が複合して第 2 層語を構成する際に、後部要素の語頭の音が変化する場合がある。たとえば、助数詞の「ホン (本)」は、「一」「六」「八」「十」に後続する際に、語頭の「ホ」が「ボ」に変わり、「三」に後続する際に、語頭の「ホ」が「バ」に変わる。このような現象を語頭音韻交替とよぶ。

語頭音韻交替は、後部要素の種類によっていくつかのパターンに類別することができる。これを用言の活用型にならって、語頭音韻交替型とよぶ。表 12 に語頭音韻交替型の一覧を示す。

表 13: 語末音韻交替型分類

語末音韻 交替型	可能な 語末の形	この型を とる語の例	語頭音韻 交替型	可能な 語頭の形	この型を とる語の例
数・一	基本形 (イチ) 促音形 (イツ) 和語形 (ヒト)	一 (イチ)	数・七	基本形 (ナナ) 漢語形 (シチ) 和オ段形 (ナノ/ナヌ)	七 (ナナ)
数・二	基本形 (ニ) 和語形 (フタ) 和ウ段形 (フツ)	二 (ニ)	数・八	基本形 (ハチ) 促音形 (ハッ) 和語形 (ヤ) 和促音形 (ヤッ) 和長音形 (ヨー)	八 (ハチ)
数・三	基本形 (サン) 和語形 (ミ) 和促音形 (ミッ)	三 (サン)	数・九	基本形 (キュー) 漢語形 (ク) 和語形 (ココノ)	九 (キュー)
数・四	基本形 (ヨン) 漢語形 (シ) 和語形 (ヨ) 和促音形 (ヨッ)	四 (ヨン)	数・十	基本形 (ジュー) 促音形 (ジュッ/ジッ) 和語形 (ト) 和長音形 (トー)	十 (ジュー)
数・五	基本形 (ゴ) 撥音形 (ゴン) 和語形 (イツ)	五 (ゴ)	数・百	基本形 (ク) 促音形 (ッ)	百 (ヒャク) 三百 (サンビャク)
数・六	基本形 (ロク) 促音形 (ロッ) 和語形 (ム) 和促音形 (ムッ) 和イ段形 (ムイ)	六 (ロク)			

8.2 語末音韻交替型

二つの第1層語が複合して第2層語を構成する際に、前部要素の語末の音が変化する場合がある。たとえば、数詞の「イチ (一)」は、助数詞「階」「泊」「本」などに接続する際に、語末の「チ」が「ッ」に変わる。このような現象を語末音韻交替とよぶ。

語末音韻交替は、前部要素の種類によっていくつかのパターンに類別することができる。これを用言の活用型にならって、語末音韻交替型とよぶ。表13に語末音韻交替型の一覧を示す。

「イチ (一)」と「ヒト (一)」などは本来、語形が異なるため、異なる索引語である。よって、音韻交替ではなく、同形異音語として扱うべきである。しかし、現状では、後続する語に応じて語形を適切に選択するといった機能を茶筌で実現するのが難しいため、これらの同形異音語も音韻交替と同じ方法で扱っている。

8.3 語頭音韻交替結合型

語頭音韻交替の際に、後部要素の語頭音がどの形をとるかは、前部要素の種類によって制約される。たとえば、数詞の「三」は、助数詞の「階」(カ濁型)、「杯」(ハ混合型)、「本」(ホ混合型)などの語頭音を濁音形に変化させ、「泊」(ハ半濁型)、「分」(フ半濁型)などの語頭音を半濁音形に変化させる。このような後部要素の語頭音韻交替形に対する制約を語頭音韻交替結合型とよぶ。表14に語頭音韻交替結合型の一覧を示す。各語

表 14: 語頭音韻交替結合型分類

語頭音韻 交替結合型	後部要素													この型を とる語の例
	カ 濁	ケ 濁	ソ 濁	ハ 半 濁	ハ 混 合	ヒ 半 濁	ヒ 混 合	フ 半 濁	ヘ 半 濁	ヘ 混 合	ホ 半 濁	ホ 混 合	ワ 混 合	
N1	基	基	基	半	半	半	半	半	半	半	半	半	基	一（イチ）
N3	濁	濁	濁	半	濁	半	濁	半	半	半/濁	半	濁	清/濁	三（サン）
N4	基	基	基	半	基	基	基	半	半	基	基	基	基	四（ヨン）
N6	基	基	基	半	半	半	半	半	半	半	半	半	清/半	六（ロク）
N8	基	基	基	半	半	半	半	半	清/半	清/半	清/半	基	基	八（ハチ）
Nj	基	基	基	半	半	半	半	半	半	半	半	半	半	十（ジュウ）
Nh	基	基	基	半	半	半	半	半	半	半	半	半	半	百（ヒャク）
Ns	基	基	基	半	基	半	濁	半	半	半	半	濁	濁	千（セン）
Nm	基	基	基	基	基	基	濁	基	半	半	半	濁	濁	万（マン）
Nn	基	基	基	半	濁	半	濁	半	半	半	半	濁	濁	何（ナン）

基: 基本形, 半: 半濁音形, 濁: 濁音形, 清/半: 清/半濁音形, 清/濁: 清/濁音形, 半/濁: 半濁/濁音形

語頭音韻交替結合型と各語頭音韻交替型とが交差するセルには、その組み合わせのときに、後部要素がどのような語頭形をとるかが記載されている。

8.4 語末音韻交替結合型

語末音韻交替の際に、前部要素の語末音がどの形をとるかは、後部要素の種類によって制約される。たとえば、助数詞の「通り」は、数詞の「八」（数・八型）、「十」（数・十型）の語末音を促音形に変化させ、「一」（数・一型）、「二」（数・二型）を和語形に変化させる。このような前部要素の語末音韻交替形に対する制約を語末音韻交替結合型とよぶ。表 15, 16 に語末音韻交替結合型の一覧を示す。各語末音韻交替結合型と各語末音韻交替型とが交差するセルには、その組み合わせのときに、前部要素がどのような語末形をとるかが記載されている。なお、表中の「-」は該当する接続が存在しないことを示す（たとえば、助数詞「つ」は「十」「百」に接続しない）。

複数の語末音韻交替結合型を持つ語があることに注意しよう。たとえば、助数詞「通り」は B1W2W8SjS 型と B1W2W3W4WjS 型の両方に分類されている。これは、「三（サン/ミ）通り」「四（ヨン/ヨ）通り」など複数の可能性があるからである。このような場合には、属性値を順序付きのリスト（コンマ区切り）で表わす。

8.5 アクセント型

アクセント型は、アクセントの位置を先頭からのモーラ数で数えることによって表わす。たとえば、頭高型の「社会」は 1 型、中高型で 3 モーラ目にアクセント核を持つ「片仮名」は 3 型である。平板型は 0 型で表わす。複数のアクセント型が可能な場合は、属性値を順序付きのリスト（コンマ区切り）で表わす（たとえば、「蠟燭」のアクセント型は“3,4”）。

表 15: 語末音韻交替結合型分類

語末音韻 交替結合型	前部要素										この型を とる語の例	
	数 ・ 一	数 ・ 二	数 ・ 三	数 ・ 四	数 ・ 五	数 ・ 六	数 ・ 七	数 ・ 八	数 ・ 九	数 ・ 十		数 ・ 百
B	基	基	基	基	基	基	基	基	基	基	基	グラム
BjS	基	基	基	基	基	基	基	基	基	促	基	ヘルツ
B1SjS	促	基	基	基	基	基	基	基	基	促	基	体（タイ）
B6SjS	基	基	基	基	基	促	基	基	基	促	基	項目（コウモク）
B8SjS	基	基	基	基	基	基	基	促	基	促	基	地点（チテン）
B1S6SjS	促	基	基	基	基	促	基	基	基	促	基	報（ホウ）
B1S8SjS	促	基	基	基	基	基	基	促	基	促	基	歳（サイ）
B6SjShS	基	基	基	基	基	促	基	基	基	促	促	羽（ワ）
B1S6S8SjS	促	基	基	基	基	促	基	促	基	促	基	区間（クカン）
B1S6SjShS	促	基	基	基	基	促	基	基	基	促	促	件（ケン）
B1S8SjShS	促	基	基	基	基	基	基	促	基	促	促	票（ヒョウ）
B6S8SjShS	基	基	基	基	基	促	基	促	基	促	促	クラス
B1S6S8SjShS	促	基	基	基	基	促	基	促	基	促	促	泊（ハク）
B7G	基	基	基	基	基	基	漢	基	基	基	基	号（ゴウ）
B9G	基	基	基	基	基	基	基	基	漢	基	基	重（ジュウ）
B4G7G9G	基	基	基	漢	基	基	漢	基	漢	基	基	月（ガツ）
B1S7GjS	促	基	基	基	基	基	漢	基	基	促	基	周年（シュウネン）
B6S7GjS	基	基	基	基	基	促	漢	基	基	促	基	カロリー
B7G8ShS	基	基	基	基	基	基	漢	促	基	基	促	羽（ワ）
B1S6S7G8SjShS	促	基	基	基	基	促	漢	促	基	促	促	遍（ヘン）
B4G5H	基	基	基	漢	撥	基	基	基	基	基	基	合（ゴウ）
B4W	基	基	基	和	基	基	基	基	基	基	基	円（エン）
B1W2W	和	和	基	基	基	基	基	基	基	基	基	駅（エキ）
B1W2W3W4W	和	和	和	和	基	基	基	基	基	基	基	駅（エキ）
B1W2W3WjW	和	和	和	基	基	基	基	基	基	和	基	袋（フクロ）
B1W2W3W4WjW	和	和	和	和	基	基	基	基	基	和	基	部屋（ヘヤ）
B1W2W3W5WjW	和	和	和	基	和	基	基	基	基	和	基	部屋（ヘヤ）
B1W2WjS	和	和	基	基	基	基	基	基	基	促	基	続き（ツヅキ）
B1W2W3WjS	和	和	和	基	基	基	基	基	基	促	基	品（シナ）
B1W2W4WjS	和	和	基	和	基	基	基	基	基	促	基	品（シナ）
B1W2W6SjS	和	和	基	基	基	促	基	基	基	促	基	株（カブ）
B1W2W8SjS	和	和	基	基	基	基	基	促	基	促	基	通り（トオリ）
B1S2WjShS	促	和	基	基	基	基	基	基	基	促	促	泊（ハク）
B1W2W3W4WjS	和	和	和	和	基	基	基	基	基	促	基	通り（トオリ）
B1W2W3W4WhS	和	和	和	和	基	基	基	基	基	基	促	振り（フリ）
B1W2W6SjShS	和	和	基	基	基	促	基	基	基	促	促	缶（カン）
B1W2W3W6SjShS	和	和	和	基	基	促	基	基	基	促	促	切れ（キレ）
B1W2W3W4W6SjShS	和	和	和	和	基	促	基	基	基	促	促	桁（ケタ）
B1W2W5W6S8SjShS	和	和	基	基	和	促	基	促	基	促	促	桁（ケタ）
B1W2W3W4W6S8SjShS	和	和	和	和	基	促	基	促	基	促	促	口（クチ）

基: 基本形, 促: 促音形, 撥: 撥音形, 漢: 漢語形, 和: 和語形

表 16: 語末音韻交替結合型分類 (続き)

語末音韻 交替結合型	前部要素											この型を とる語の例
	数 ・ 一	数 ・ 二	数 ・ 三	数 ・ 四	数 ・ 五	数 ・ 六	数 ・ 七	数 ・ 八	数 ・ 九	数 ・ 十	数 ・ 百	
B4W9G	基	基	基	和	基	基	基	基	漢	基	基	日 (ニチ)
B4W7G9G	基	基	基	和	基	基	漢	基	漢	基	基	時 (ジ)
B1W2W3W4W7G	和	和	和	和	基	基	漢	基	基	基	基	幕 (マク)
B1W2W4W7G9G	基	基	基	和	基	基	漢	基	漢	基	基	人 (ニン)
W	和	和	和	和	和	和	基	和	和	—	—	つ
W9B	和	和	和	和	和	和	基	和	基	和	基	柱 (ハシラ)
W9BjS	和	和	和	和	和	和	基	和	基	促	基	度 (タビ)
W9BhS	和	和	和	和	和	和	基	和	基	和	促	切れ (キレ)
W6B8S9B	和	和	和	和	和	基	基	促	基	和	基	粒 (ツブ)
W6B8B9BjS	和	和	和	和	和	基	基	基	基	促	基	皿 (サラ)
W6S8B9BjS	和	和	和	和	和	促	基	基	基	促	基	株 (カブ)
W6S8B9BhS	和	和	和	和	和	促	基	基	基	和	促	組 (クミ)
W5B9G	和	和	和	和	基	和	基	和	漢	和	基	度 (タビ)
W3D4D6D8D	和	和	和促	和促	和	和促	基	和促	和	—	—	つ
D	—	和ウ	和促	和促	和	和イ	和オ	和長	和	和長	—	日 (カ)

基: 基本形, 促: 促音形, 漢: 漢語形, 和: 和語形,
和促: 和促音形, 和長: 和長音形, 和イ: 和イ段形, 和ウ: 和ウ段形, 和オ: 和オ段形

8.6 アクセント結合型

二つの第1層語が複合して第2層語を構成する際に、前後の要素のもともとのアクセントの位置が移動する場合がある。たとえば、「社会」は1型、「生活」は0型であるが、これらが複合して「社会生活」になると、4型になる。同様に、「社会」と接尾辞の「系」が複合して「社会系」になると、0型になる。このようなアクセント変化はかなり規則的であり、変化規則は概ね後部要素の性質によって決まる（接頭辞と名詞が複合する際のアクセント変化規則は前部要素である接頭辞の性質によって決まる）。このような複合語形成の際のアクセント変化に対する制約をアクセント結合型とよぶ。

同様のことが、自立語と付属語が接続して文節を構成する際にも生じる。たとえば、アクセント型が1型の「社会」が格助詞「さえ」と接続して「社会さえ」になると、アクセント型は1型のままであるが、0型の「生活」が「さえ」と接続して「生活さえ」になると、5型になる。このように助詞・助動詞が文節を形成する際にアクセント変化に与える制約も、アクセント結合型とよぶ。

アクセント結合型は、匂坂・佐藤(1983)のものを拡張して設計した。

8.6.1 名詞・接尾辞

普通名詞や名詞性接尾辞が複合語の後部要素になる際のアクセント結合型の一覧を表17に示す。複合語のアクセント型の欄には、各アクセント結合型について、複合語のアクセント型を決定する規則を記載している。たとえば、C1型では、前部要素のモーラ数(N_1)と後部要素のアクセント型(M_2)の和が複合語のアクセ

表 17: アクセント結合型分類 (普通名詞・接尾辞)

アクセント結合型	複合語のアクセント型	この型をとる語の例
C1	$N_1 + M_2$	手続き (テツヅキ), 日間 (ニチカン)
C2	$N_1 + 1$	生活 (セイカツ), 時間 (ジカン)
C3	N_1	湾 (ワン), 学 (ガク)
C4	0	島 (シマ), 系 (ケイ)
C5	M_1	殿 (ドノ)
C10	$M_1 M_2$	等々 (トウトウ)

N_1 : 前部要素のモーラ数, M_1 : 前部要素のアクセント型, M_2 : 後部要素のアクセント型

表 18: アクセント結合型分類 (接頭辞)

アクセント結合型	文節のアクセント型		この型をとる語の例
	後部要素が平板・尾高型	後部要素が頭高・中高型	
P1	0	$N_1 + M_2$	御 (ゴ)
P2	$N_1 + 1$	$N_1 + M_2$	総 (ソウ)
P13	M_1 または $M_1 M_2$		現 (ゲン)
P4	$N_1 + 1$ または M_1	M_1 または $N_1 + M_2$ または $M_1 M_2$	両 (リョウ)
P6	0		再来 (サライ)
P14	M_1	$N_1 + M_2$	要 (ヨウ)

N_1 : 前部要素のモーラ数, M_1 : 前部要素のアクセント型, M_2 : 後部要素のアクセント型

ント型になる (「訴訟 | 手続き」→ $N_1 + M_2 = 3 + 2 = 5$)。ただし、一般的な原則として、規則から求まるアクセントの位置が重音節の二拍目 (促音・撥音・長音・二重母音の後部) や無声化母音である場合には、左に 1 モーラずれる (「情報 | 学」→ $N_1 = 4$ が長音なので 3)。

一般に、3 モーラ以上の語については、平板型と尾高型 (最終音節にアクセントがある型) の語は C2 型に、それ以外は C1 型に分類される。ただし、例外も多くある。2 モーラ以下の語については、一般に、C3 型か C4 型に分類されるが、これといった規則性はない。また、2 モーラの語であっても C1 型や C2 型に分類される場合がある (「か所」→ C1 型、「点」→ C2 型)。

C5 型や C10 型は特殊である。とくに、C10 型では、前部要素と後部要素が韻律的には複合せず、それぞれのアクセント型がそのまま残る (複合語アクセント型の “ $M_1 | M_2$ ” は、アクセント単位が二つに分かれ、それぞれのアクセント型が M_1 と M_2 であることを表す)。

8.6.2 接頭辞

名詞・形状詞接続の接頭辞が複合語の前部要素になる際のアクセント結合型の一覧を表 18 に示す。接頭辞と名詞の複合の場合は、一般に、後部要素が平板型や尾高型 (最終音節にアクセントがある型) であるかどうかによって、適用される規則が異なる。たとえば、P1 型では、後部要素が平板・尾高型の場合は複合語のアクセント型は 0 型であり (「御 | 結婚」→ 0)、頭高・中高型の場合は前部要素のモーラ数 (N_1) と後部要素のアクセント型 (M_2) の和が複合語のアクセント型になる (「御 | 利益」→ $N_1 + M_2 = 1 + 1 = 2$)。

表 19: アクセント結合型分類（助詞・助動詞）

アクセント 結合型	文節のアクセント型		この型を とる語の例
	前部要素が 平板型	前部要素が 起伏型	
F1	M_1		が（格助詞）, た（助動詞）
F2@ M	$N_1 + M$	M_1	か（終助詞）, です（助動詞）
F3@ M	M_1	$N_1 + M$	せる（助動詞）
F4@ M	$N_1 + M$		ます（助動詞）
F5	0		だけ（副助詞）
F6@ M, L	$N_1 + M$	$N_1 + L$	ず（助動詞）

N_1 : 前部要素のモーラ数, M_1 : 前部要素のアクセント型, M, L : 後部要素の結合アクセント型

P13 型や P4 型では、複数の複合語アクセント型が可能である。これらは、構文・意味上の相違などによって使い分けられる（「(東西) 両大関」→ $N_1 + 1 = 2 + 1 = 3$ 、「(魁皇・千代大海の) 両大関」→ $M_1 | M_2 = 1 | 1$ ）。

8.6.3 助詞・助動詞

助詞・助動詞（「形状詞-助動詞語幹」を含む）が自立語と接続して文節を作る際のアクセント結合型の一覧を表 19 に示す。助詞・助動詞が自立語と接続して文節を作る場合は、一般に、前部要素が平板型か起伏型かによって、適用される規則が異なる。たとえば、F3 型では、前部要素が平板型の場合は文節のアクセント型は前部要素のアクセント型 (M_1) によって決まり（「行わ|せる」→ $M_1 = 0$ ）、起伏型の場合には前部要素のモーラ数 (N_1) と後部要素の結合アクセント価 (M) の和が文節のアクセント型になる（「手伝わ|せる」→ $N_1 + M = 4 + 1 = 5$ ）。助詞・助動詞では、単独発声した場合のアクセント型とは別に、接続に伴って生じるアクセント核の位置を指定する結合アクセント価が必要となる。“F3@ M ” などの M がこれを表わす。

さらに、助詞・助動詞では、接続する前部要素の品詞（名詞・動詞・形容詞）によって、文節のアクセント型が異なる場合がある。よって、一般に、アクセント結合型を「品詞 % 当該品詞に対するアクセント結合型」の列によって表わす。たとえば、「だけ」のアクセント結合型は、“名詞 %F5, 動詞 %F1, 形容詞 %F1” となる。

参考文献

- 浅原正幸・松本裕治. (2003). ipadic version 2.6.3 ユーザーズマニュアル.
- 伝康晴・宇津呂武仁・山田篤・浅原正幸・松本裕治. (2002). 話し言葉研究に適した電子化辞書の設計. 第2回「話し言葉の科学と工学」ワークショップ講演予稿集 (pp. 39–46). 東京.
- 小林哲則・板橋秀一・速水悟・竹澤寿幸. (1992). 日本音響学会研究用連続音声データベース. 日本音響学会誌, 48, 888–893.
- 窪園晴夫. (1995). 語形成と音韻構造. 東京: くろしお出版.
- 前川喜久雄・籠宮隆之・小磯花絵・小椋秀樹・菊池英明. (2000). 日本語話し言葉コーパスの設計. 音声研究, 4(2), 51–61.
- 前川喜久雄・菊池英明・籠宮隆之・山口昌也・小磯花絵・小椋秀樹. (2001). 『日本語話し言葉コーパス』の構築における計算機利用. 日本語学, 20(13), 61–79.
- 松本裕治. (2000). 形態素解析システム『茶釜』. 情報処理, 41, 1208–1214.
- 小椋秀樹. (2001). 話し言葉コーパスの単位認定基準について. 「話し言葉の科学と工学」ワークショップ講演予稿集 (pp. 21–28). 東京.
- 新情報処理開発機構 (RWCP) テキスト・サブ・ワーキンググループ. (1998). 研究開発用知的資源: タグ付きテキストコーパス 報告書.
- 匂坂芳典・佐藤大和. (1983). 日本語単語連鎖のアクセント規則. 電子情報通信学会論文誌, J66-D, 849–856.

付録 A UniDic を用いた形態素解析

UniDic を自動形態素解析で用いる際の注意点について述べる。

4.2.2 節で述べたように、UniDic では、

1. 語彙素: 〈語概念〉
2. 索引語: 語彙素を定義する属性群 + 〈語形〉
3. (a) 書字語: 索引語を定義する属性群 + 〈書字形〉
(b) 発音語: 索引語を定義する属性群 + 〈発音形〉

という階層によって、語を定義している。形態素解析の目的は、入力文中の語を同定することであるから、形態素解析システムは、与えられた入力に対して、これらの属性を推定しなければならない。たとえば、以下のようである。

入力 ウィンドウズを仮名で書く

出力 〈語概念 = ウィンドウズ!名詞-固有名詞-一般, 語形 = ウィンドウズ,
書字形 = ウィンドウズ, 発音形 = ウィンドーズ〉
〈語概念 = ヲ!を!助詞-格助詞-一般, 語形 = ヲ, 書字形 = を, 発音形 = オ〉
〈語概念 = カナ!仮名!名詞-普通名詞-一般, 語形 = カナ, 書字形 = 仮名, 発音形 = カナ〉
〈語概念 = デ!で!助詞-格助詞-一般, 語形 = デ, 書字形 = で, 発音形 = デ〉
〈語概念 = カ*ク!書く!動詞-一般, 語形 = カク, 書字形 = 書く, 発音形 = カク〉

しかし、多くの形態素解析システムの出力は以下のようなものである。

出力 〈書字形 = ウィンドウズ, 品詞 = 名詞-固有名詞-一般〉
〈書字形 = を, 品詞 = 助詞-格助詞-一般〉
〈書字形 = 仮名, 品詞 = 名詞-普通名詞-一般〉
〈書字形 = で, 品詞 = 助詞-格助詞-一般〉
〈書字形 = 書く, 品詞 = 動詞-一般, 活用型 = 五段-カ行-一般-一般〉

これは上記のものとは等価ではない。なぜなら、「仮名」という書字形を持つ語は「カナ」と「カメイ」の二つがあり、その判別がなされていないからである。

このように、形態素解析システムにおいては、形態素解析という問題を「入力文を部分文字列に分割し、それらに品詞・活用型を割り当てること」ととらえており、これは本来の形態素解析の一部でしかない。UniDic を用いた茶筌では、語概念・語形・書字形・発音形をすべて推定することができるが、その基礎となる解析モデルは、やはり書字形・品詞・活用型のみにもとづいたものであり、語概念・語形・発音形の推定は必ずしも精度が高いとはいえない。よって、今後、解析モデルの拡張、もしくは、後処理による語形・発音選択によって、この点を改良していく必要があるだろう。

さて、上のような問題のとらえ方によって、形態素解析システムで用いられる辞書は、国語辞典とは異なり、書字形を見出しにして設計されている。たとえば、UniDic 茶筌辞書でも、2.4 節で述べたように、見出し語は書字形である。よって、書字形が同じ語は、異なる語であっても、以下のように、同一見出しのもとにまとめて記述されている。

(品詞 (名詞 普通名詞 一般))

((見出し語 (仮名 2521)) (読み {カナ/カメイ}) (発音 {カナ/カメー})

(付加情報 '{lex="カナ!仮名" indexForm="カナ" indexOrth="仮名"

aType="0" aConType="C3"

/

lex="カメイ!仮名" indexForm="カメイ" indexOrth="仮名"

aType="0" aConType="C2"}')

読み・発音・付加情報のバリエーションは、一般に、「{選択肢 1/選択肢 2/.../選択肢 N}」の形で選言によって記述されており、読み・発音・付加情報の選択肢の順序は一致する。

このように、UniDic の語の定義の階層性と、UniDic 茶筌辞書における記述の階層性とは、一致しないので注意が必要である。

付録 B 他の品詞体系との違い

B.1 国語研短単位体系との違い

- 短単位と UniDic 第 1 層の単位はほぼ一致する。
- 短単位体系では、活用形に意志推量形を設ける代わりに、意志・推量の助動詞「う」「よう」を立てている (単位の一一致の例外)。
- 短単位体系には、「その他」に相当する補助記号類がなく、「・」などは直前の語にくっつけている (単位の一一致の例外)。
- 短単位体系の品詞は概ね UniDic の中分類まで、活用形は概ね UniDic の行分類までである。
- 短単位体系には、人名・地名以外の固有名詞はない。
- 短単位体系では、UniDic の「名詞-形状詞可能」「名詞-副詞可能」を、使用された文脈に応じて「名詞」と「形状詞」「副詞」に分類している。
- 短単位体系では、UniDic の「基本形」を、使用された文脈に応じて「終止形」と「連体形」に分類している。

B.2 IPA 体系との違い

- IPA 体系の単位は斉一性がなく、UniDic 第 1 層から第 3 層 (あるいはそれより長い単位) が混在している。
- IPA 体系では、活用形に意志推量形を設ける代わりに、意志・推量の助動詞「う」を立てている (「よう」の「よ」は活用語尾に含めている)。
- IPA 体系の品詞は概ね UniDic の品詞と一致する。活用形は概ね UniDic の行分類までである。
- IPA 体系では、UniDic の「名詞-形状詞可能」と「形状詞」の区別がなく、いずれも「名詞-形容動詞語幹」としている。
- IPA 体系では、接尾辞は「名詞」「動詞」「形容詞」の下位分類にしている。
- IPA 体系では、「浅間山」「血液型」の「ヤマ (山)」「ガタ (型)」を接尾辞にするなど、接尾辞の認定の仕方が UniDic とは異なる (UniDic では、いずれも普通名詞)。

付録 C To Do リスト

- 語彙の拡充。RWCP コーパスを UniDic 体系に移植するなどして、語彙を拡充する。
- 語彙素への意味属性の導入。分類語彙表などの利用を考える。
- 索引語への語種属性の導入。漢語・和語・外来語などを分類する。
- 人名（姓・名）で東洋名と西洋名を区別する。姓と名の順序が逆になるため。
- 「名詞-数詞-アラビア数字」の廃止。アラビア数字は位取りして解析できないため、位取りすべき（であると推定される）数字はすべて前処理で漢数字に変換することにし、アラビア数字が出現した場合は、数詞の単独用法として扱うことにする。
- 「イチ（一）」「ヒト（一）」などを索引語のレベルで区別（現状では、音韻交替として扱っている）。文脈に依存した語形選択を実現し、本来、語形の違いとしてとらえるべきものを音韻交替の対象外とすることで、音韻交替属性を整理する。
- 音韻交替属性の拡張。数詞がらみ以外の一般の連濁や語末音変化を扱えるように拡張する。
- アクセント属性の整理と拡張。アクセント結合型を整理する。複合動詞や複合形容詞のアクセント変化を扱えるように拡張する。
- エラーの修正