

ipadic version 2.7.0 ユーザーズマニュアル

浅原正幸 松本裕治

平成 15 年 11 月

IPADIC version 2.7.0 Users Manual

Masayuki Asahara and Yuji Matsumoto

Copyright (c) 2003 Nara Institute of Science and Technology, All rights reserved.

This edition is for "IPADIC for Japanese" version 2.7.0.

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 IPADIC version 2.7.0 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 1.0b	25 May 1998
version 1.0	27 April 1999
version 2.0	15 December 1999
version 2.1	30 December 1999
version 2.4.0	6 December 2000
version 2.5.0	13 April 2001
version 2.6.0	19 June 2003
version 2.7.0	15 November 2003

目次

1	インストール	2
1.1	UNIX における辞書インストール	2
1.2	Windows における辞書インストール	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品詞定義	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	4
2.4	辞書ファイルのフォーマット	4
2.5	接続表ファイルのフォーマット	7
3	chasenrc ファイル	8
4	形態素エントリの追加	14
4.1	各ファイルの編集	14
4.2	システム辞書の再コンパイル (UNIX 版)	15
4.3	ユーザー辞書の再コンパイル (UNIX 版)	15
4.4	辞書の再コンパイル (Windows 版)	16
5	IPA 品詞体系	16
5.1	名詞	17
5.1.1	名詞-一般	17
5.1.2	名詞-固有名詞-一般	17
5.1.3	名詞-固有名詞-人名-一般	17
5.1.4	名詞-固有名詞-人名-姓	17
5.1.5	名詞-固有名詞-人名-名	18
5.1.6	名詞-固有名詞-組織	18
5.1.7	名詞-固有名詞-地域-一般	18
5.1.8	名詞-固有名詞-地域-国	18
5.1.9	名詞-代名詞-一般	18
5.1.10	名詞-代名詞-縮約	18
5.1.11	名詞-副詞可能	18
5.1.12	名詞-サ変接続	19
5.1.13	名詞-形容動詞語幹	19
5.1.14	名詞-ナイ形容詞語幹	19
5.1.15	名詞-数	19
5.1.16	名詞-非自立-一般	19
5.1.17	名詞-非自立-副詞可能	20
5.1.18	名詞-非自立-助動詞語幹	20
5.1.19	名詞-非自立-形容動詞語幹	20
5.1.20	名詞-特殊-助動詞語幹	20
5.1.21	名詞-接尾-一般	20
5.1.22	名詞-接尾-人名	21
5.1.23	名詞-接尾-地域	21
5.1.24	名詞-接尾-サ変接続	21

5.1.25	名詞-接尾-助動詞語幹	21
5.1.26	名詞-接尾-形容動詞語幹	21
5.1.27	名詞-接尾-副詞可能	21
5.1.28	名詞-接尾-助数詞	21
5.1.29	名詞-接尾-特殊	22
5.1.30	名詞-接続詞的	22
5.1.31	名詞-動詞非自立的	22
5.2	接頭詞	22
5.2.1	接頭詞-名詞接続	22
5.2.2	接頭詞-数接続	22
5.2.3	接頭詞-動詞接続	22
5.2.4	接頭詞-形容詞接続	22
5.3	動詞	23
5.3.1	動詞-自立 力変【活用形】	24
5.3.2	動詞-非自立 力変【活用形】	24
5.3.3	動詞-自立 サ変・スル【活用形】	24
5.3.4	動詞-自立 サ変・-スル【活用形】	25
5.3.5	動詞-自立 サ変・-ズル【活用形】	25
5.3.6	動詞-自立 一段【活用形】	25
5.3.7	動詞-非自立 一段【活用形】	25
5.3.8	動詞-接尾 一段【活用形】	25
5.3.9	動詞-自立 五段・カ行イ音便【活用形】	25
5.3.10	動詞-非自立 五段・カ行イ音便【活用形】	26
5.3.11	動詞-非自立 五段・カ行促音便【活用形】	26
5.3.12	動詞-非自立 五段・カ行促音便【活用形】	26
5.3.13	動詞-自立 五段・ガ行【活用形】	26
5.3.14	動詞-自立 五段・サ行【活用形】	26
5.3.15	動詞-非自立 五段・サ行【活用形】	26
5.3.16	動詞-自立 五段・タ行【活用形】	26
5.3.17	動詞-自立 五段・ナ行【活用形】	26
5.3.18	動詞-自立 五段・バ行【活用形】	27
5.3.19	動詞-自立 五段・マ行【活用形】	27
5.3.20	動詞-非自立 五段・マ行【活用形】	27
5.3.21	動詞-自立 五段・ラ行【活用形】	27
5.3.22	動詞-非自立 五段・ラ行【活用形】	27
5.3.23	動詞-接尾 五段・ラ行【活用形】	27
5.3.24	動詞-自立 五段・ラ行特殊【活用形】	27
5.3.25	動詞-非自立 五段・ラ行特殊【活用形】	27
5.3.26	動詞-自立 五段・ワ行ウ音便【活用形】	27
5.3.27	動詞-非自立 五段・ワ行ウ音便【活用形】	28
5.3.28	動詞-自立 五段・ワ行促音便【活用形】	28
5.3.29	動詞-非自立 五段・ワ行促音便【活用形】	28
5.3.30	動詞-自立 四段・ハ行【活用形】	28
5.3.31	動詞-自立 ラ変【活用形】	28
5.3.32	動詞-自立 上二・ハ行【活用形】	28
5.3.33	動詞-自立 下二・ア行【活用形】	29

5.3.34	動詞-自立 一段・得ル【活用形】	29
5.4	形容詞	29
5.4.1	形容詞-自立 形容詞・アウオ段【活用形】	31
5.4.2	形容詞-自立 形容詞・イ段【活用形】	31
5.4.3	形容詞-自立 形容詞・イイ【活用形】	31
5.4.4	形容詞-自立 形容詞・不変化型【活用形】	31
5.4.5	形容詞-非自立 形容詞・アウオ段【活用形】	31
5.4.6	形容詞-非自立 形容詞・イ段【活用形】	31
5.4.7	形容詞-非自立 形容詞・イイ【活用形】	31
5.4.8	形容詞-非自立 形容詞・不変化型【活用形】	31
5.4.9	形容詞-接尾 形容詞・アウオ段【活用形】	32
5.4.10	形容詞-接尾 形容詞・イ段【活用形】	32
5.5	副詞	32
5.5.1	副詞-一般	32
5.5.2	副詞-助詞類接続	32
5.6	連体詞	32
5.6.1	連体詞	32
5.7	接続詞	32
5.7.1	接続詞	32
5.8	助詞	32
5.8.1	助詞-格助詞-一般	32
5.8.2	助詞-格助詞-引用	33
5.8.3	助詞-格助詞-連語	33
5.8.4	助詞-接続助詞	33
5.8.5	助詞-係助詞	33
5.8.6	助詞-副助詞	33
5.8.7	助詞-並立助詞	34
5.8.8	助詞-終助詞	34
5.8.9	助詞-副助詞 / 並立助詞 / 終助詞	34
5.8.10	助詞-連体化	34
5.8.11	助詞-副詞化	34
5.8.12	助詞-特殊	34
5.8.13	助詞-間投助詞	35
5.9	助動詞	35
5.9.1	助動詞 五段・ラ行アル【活用形】	35
5.9.2	助動詞 五段・ラ行ゴザル【活用形】	35
5.9.3	助動詞 形容詞・イ段【活用形】	35
5.9.4	助動詞 特殊・ナイ【活用形】	35
5.9.5	助動詞 特殊・タ【活用形】	35
5.9.6	助動詞 特殊・ダ【活用形】	35
5.9.7	助動詞 特殊・デス【活用形】	35
5.9.8	助動詞 特殊・ドス【活用形】	36
5.9.9	助動詞 特殊・ジャ【活用形】	36
5.9.10	助動詞 特殊・マス【活用形】	36
5.9.11	助動詞 特殊・ヌ【活用形】	36
5.9.12	助動詞 特殊・ヤ【活用形】	36

5.9.13 助動詞 不変化型【活用形】	36
5.9.14 助動詞 文語・??【活用形】	36
5.10 感動詞	37
5.10.1 感動詞	37
5.11 記号	37
5.11.1 記号-一般	37
5.11.2 記号-アルファベット	37
5.11.3 記号-句点	37
5.11.4 記号-読点	37
5.11.5 記号-空白	37
5.11.6 記号-括弧開	37
5.11.7 記号-括弧閉	37
5.12 フィラー	38
5.12.1 フィラー	38
5.13 その他	38
5.13.1 その他-間投	38
参考文献	38
付録	40
A 変更点	40
A.1 ipadic-2.6.3 から ipadic-2.7.0 への変更点	40
A.2 ipadic-2.6.2 から ipadic-2.6.3 への変更点	40
A.3 ipadic-2.6.1 から ipadic-2.6.2 への変更点	40
A.4 ipadic-2.6.0 から ipadic-2.6.1 への変更点	40
A.5 ipadic-2.5.1 から ipadic-2.6.0 への変更点	40
A.6 ipadic-2.5.0 から ipadic-2.5.1 への変更点	41
A.7 ipadic-2.4.X から ipadic-2.5.0 への変更点	41
B 辞書の著作権	41

はじめに

形態素解析システム茶筌は、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 (version2.3.2 以降) 用の日本語辞書 (ipadic2.7.0) について説明したものです。本辞書では、情報処理振興事業協会 (IPA) で設定された IPA 品詞体系 (THiMCO97) に基づいて一部修正を加えました。本説明書は新情報処理開発機構 (RWCP) による「テキストデータベース報告書 (平成 8 年度)」に掲載された IPA 品詞体系 (THiMCO97) の説明を許可を得て抜粋し、一部修正を施したものです。

なお、現在の IPA 品詞体系日本語辞書は、1998 年 5 月に公開した IPA 品詞体系日本語辞書 (ipadic1.0b2) に対して、奈良先端科学技術大学院大学情報科学研究科鹿野清宏教授を代表とする「日本語ディクテーション基本ソフトウェアの開発」(IPA 独創的先進的情報技術に関わる研究開発) のグループの方々に大幅な修正、改良を行っていただき、その後、様々な修正を行ったものです。

本辞書システム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。

本マニュアルに関するお問い合わせは以下にお願いします。

〒 630-0192

奈良県生駒市高山町 8916-5

奈良先端科学技術大学院大学

情報科学研究科 自然言語処理学講座

Tel: (0743)72-5240, Fax: (0743)72-5249

E-mail: chasen@is.aist-nara.ac.jp

また以下の URL にて最新情報を提供しています。

URL: <http://chasen.aist-nara.ac.jp/>

1 インストール

1.1 UNIX における辞書インストール

本辞書は chasen-2.3.2 以上を必要とする．先に茶筌をダウンロードしインストールする．
通常のインストール方法

1. ./configure を実行する．

```
%. /configure
```

インストール先ディレクトリは茶筌側が要求するため自動的に設定される．インストール先を手動で変更したい場合は以下のように実行する．

```
% ./configure --with-dicdir=/home/masayu-a
```

すると，/home/masayu-a/ipadic 以下に辞書が作成される．

2. make を実行する．

```
% make
```

OS 標準の make を使うとコンパイルに失敗することがある．その場合は GNU make を使用する．

3. root 権限で make install を実行する．

```
# make install
```

デフォルトでは /usr/local/share/chasen/dic/ipadic にインストールされる (環境によって多少異なることもある)．

ユーザーのディレクトリにインストールする場合には，root 権限は必要としない．

4. /usr/local/etc/chasenc の編集

chasen を初めてインストールし，辞書を初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/chasenc を作成する．そうでない場合，パッケージ中に生成される chasenc を見ながら /usr/local/etc/chasenc を編集する必要がある．

1.2 Windows における辞書インストール

以下の説明では，wincha が次に示すような形でインストールされていると仮定する．

```
c:\Program Files\chasen21\dic  
c:\Program Files\chasen21\dll  
c:\Program Files\chasen21\doc  
c:\Program Files\chasen21\mkchadic  
c:\Program Files\chasen21\wincha  
c:\Program Files\chasen21\wvshell
```


WinCha をインストールすると、自動的に IPADIC がインストールされる。しかし、ユーザーが辞書に手を入れる場合には、SJIS の辞書パッケージを用意する必要がある。SJIS の辞書は以下にある。

<http://chasen.aist-nara.ac.jp/stable/ipadic/win/>

展開された辞書ファイル (Noun.dic など、拡張子 .dic で終わるファイル)、接続表ファイル (connect.cha)、品詞定義ファイル (grammar.cha)、活用型定義ファイル (ctypes.cha)、活用形定義ファイル (cforms.cha) を、Windows 版の茶筌がインストールされているディレクトリの中にある c:\Program Files\chasen21\dic というディレクトリの下にコピーする。

辞書パッケージ中にある、Makefile.bat を c:\Program Files\chasen21 にコピーし、コマンドプロンプトで Makefile.bat を実行する。

```
C:\Program Files\chasen21> Makefile.bat
```

Windows XP/2000/NT などでは Administrator 権限でインストールする必要がある。

2 各ファイルのフォーマット

2.1 品詞定義ファイルの品詞定義

フォーマットファイル grammar.cha には品詞のリストを記述する。活用する品詞である場合 (最上位品詞定義) に % がつけられる。活用する品詞には、ctypes.cha に可能な活用型を列挙し、cforms.cha に可能な活用形を記述する必要がある。

```
(接頭詞
  (名詞接続)
  (動詞接続)
  (形容詞接続)
  (数接続))

(動詞%
  (自立)
  (非自立)
  (接尾))
```

- <品詞定義> ::= ”(<最上位品詞情報> (<下位品詞情報>)*)”
- <最上位品詞定義> ::= <最上位品詞名> | ”<最上位品詞名>%”
- <下位品詞定義> ::= <品詞細分類名> | ”<品詞細分類名> (<下位品詞情報>)*”

2.2 活用型定義ファイルのフォーマット

活用型定義ファイル ctypes.cha には、各品詞がどのような活用型を取りうるかを記述する。

```
( (形容詞 自立)
  (形容詞・アウオ段
   形容詞・イ段
   不変化型)
)
```

- 〈活用型定義〉 ::= ”(((品詞名)) ((活用型)*))”

2.3 活用形定義ファイルのフォーマット

活用形定義ファイル cforms.cha には、各活用型がどのような活用形を取り、また各活用形がどのような活用語尾を取りうるかを記述する。活用語尾は、漢字形、かな形、発音形を記述することができる。

```
(形容詞・イ段
 ( ; (語幹 * )
 (基本形 い イ )
 (文語基本形 * * )
 (未然又接続 から カラ )
 (未然ウ接続 かろ カロ )
 (連用夕接続 かつ カツ )
 (連用テ接続 く ク )
 (連用テ接続 くっ クツ )
 (連用ゴザイ接続 ゆう ユウ ユー)
 (連用ゴザイ接続 ゆう ユウ ユー)
 (体言接続 き キ )
 (仮定形 けれ ケレ )
 (命令e かれ カレ )
 (仮定縮約1 けりゃ ケリャ)
 (仮定縮約2 きゃ キャ )
 (ガル接続 * ))
)
```

- 〈活用形定義〉 ::= ”(〈活用型名〉 (〈活用形情報〉*))”
- 〈活用形情報〉 ::= ”(〈活用形名〉 〈漢字形活用語尾〉 〈かな形活用語尾〉 〈発音活用語尾〉)” | ”(〈活用形名〉 〈漢字形活用語尾〉 〈かな形活用語尾〉)” | ”(〈活用形名〉 〈漢字形活用語尾〉)”

2.4 辞書ファイルのフォーマット

以下は辞書ファイルの例である。辞書ファイルは各品詞毎に分割されている。

(品詞 (名詞 一般)) ((見出し語 (お正月 3641)) (読み オショウガツ) (発音 オショウガツ))

(品詞 (動詞 自立)) ((見出し語 (あきらめる 2377)) (読み アキラメル) (活用型 一段))

(品詞 (名詞 一般)) ((見出し語 (天文学 3556)) (読み テンモンガク))

(複合語

((品詞 (名詞 一般)) (見出し語 天文) (読み テンモン))

((品詞 (名詞 接尾 一般)) (見出し語 学) (読み ガク)))

辞書ファイル中の形態素エントリの定義は以下の通り .

- <形態素エントリ> ::= "((<品詞情報>) (<見出し語情報>) <形態素情報>)*"
- <品詞情報> ::= "(品詞 (<品詞名>))"
- <見出し語情報> ::= "(見出し語 (<見出し語>) <形態素生起コスト>)" | "(見出し語 <見出し語>)"
- <形態素情報> ::= <読み情報> | <発音情報> | <活用型情報> | <付加情報> | <意味情報> | <複合語情報>
- <読み情報> ::= "(読み <読み>)"
- <発音情報> ::= "(発音 <発音>)"
- <活用型情報> ::= "(活用型 <活用型>)"
- <複合語情報> ::= "(複合語 <構成語エントリ>)*"
- <構成語エントリ> ::= "((<品詞情報>) <見出し語情報> <構成語形態素情報>)*"
- <構成語形態素情報> ::= <読み情報> | <発音情報> | <活用型情報> | <付加情報> | <意味情報> | <活用形情報>
- <活用形情報> ::= "(活用形 <活用形>)"

尚, <形態素情報> および <構成語形態素情報> の中に現れる項目の重複は許されない .

- <品詞名>

品詞名 . 階層構造をもつ . 各階層は空白で区切られる .

例:

(品詞 (名詞 一般))

(品詞 (動詞 自立))

(品詞 (名詞 接尾 一般))

- <見出し語>

文章中出现する表記 . 活用語は基本形のみを登録する .

例:

(見出し語 (お正月 3641))
(見出し語 (あきらめる 2377))
(見出し語 (天文学 3556))
(見出し語 天文)
(見出し語 学)

- 〈形態素生起コスト〉

見出し語の横にある数字を形態素生起コストという。数値が小さいほど出現しやすい語という意味になる。IPADIC の形態素生起コストは解析済みのデータから学習した単語の出現確率を元に計算した数値が付与されている。

ユーザーが形態素を追加する際には、近い頻度を持つ形態素と同じ程度の数値にしておけば、ほとんどの場合素直に解析に反映されるはずである。反映されない場合は、形態素コストの値を小さくしてみるとよい。

例:

(見出し語 (お正月 3641))
(見出し語 (あきらめる 2377))
(見出し語 (天文学 3556))

- 〈読み〉

かな表記。カタカナで記述される。

例:

(読み オショウガツ)
(読み アキラメル)
(読み テンモンガク)
(読み テンモン)
(読み ガク)

- 〈発音〉

発音表記。カタカナで記述される。

例:

(発音 オショーガツ)

- 〈活用型〉

活用語は「活用型」を要求する。活用型は ctypes.cha に定義されたもののみ許される。

例:

(活用型 五段・サ行)

- 〈活用形〉

複合語を構成する形態素が活用語でありかつ基本形でない場合に、構成語エントリを記述するために利用する。

例:

(活用形 未然ウ接続)

- 〈付加情報〉

単語エントリに対する付加情報。ユーザーが自由に利用できる。アクセント情報や他の品詞体系での品詞名などが記述できる。

例:

(付加情報 アクセント型=4)

- 〈意味情報〉

単語エントリに対する意味情報。ユーザーが自由に利用できる。シソーラスの情報や国語辞書の項目などを記述することができる。

例:

(意味情報 "思い切る。仕方がないと断念する。")

2.5 接続表ファイルのフォーマット

以下は接続表ファイル connect.cha 中の接続規則の例である。*は適合する全可能性を示す。ファイルの前方にある規則に適合する接続規則は、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。

```
(( (( (名詞 固有名詞 人名 姓) ))
  (( (名詞 接尾 人名) )) ) 842)

(((( (動詞 自立) 五段・ラ行アル 連用形 ))
  (( (助動詞) 特殊・マス )) ) 604)

(((( (助詞 接続助詞) * * て))
  (( (助詞 係助詞) * * も))
  (( (形容詞 非自立) 形容詞・アウオ段 * よい)) ) 35)
```

- 〈接続規則エントリ〉 ::= ”(〈接続情報〉 〈接続コスト〉)”

- 〈接続情報〉 ::= 〈品詞定義〉 〈品詞定義〉+

- 〈品詞定義〉+ ::= 〈品詞定義〉 | 〈品詞定義〉+

- 〈品詞定義〉 ::= ”(〈品詞情報〉 〈活用型情報〉 〈活用形情報〉 〈語彙化品詞規則〉)” | ”(〈品詞情報〉 〈活用型情報〉 〈活用形情報〉)” | ”(〈品詞情報〉 〈活用型情報〉)” | ”(〈品詞情報〉)”

- 〈品詞情報〉 ::= ”(〈品詞名〉)”
- 〈活用型情報〉 ::= ”〈活用型〉” | ”*”
- 〈活用形情報〉 ::= ”〈活用形〉” | ”*”
- 〈語彙化品詞規則〉 ::= ”〈語彙化品詞定義〉” | ”*”

3 chasenrc ファイル

chasenrc ファイルは形態素解析プログラムに必要な様々な選択肢を定義するために用いられる。これらの定義は通常、PREFIX/etc/chasenrc に記述されるが、利用者のホームディレクトリの '.chasenrc' というファイルに記述することもできる。起動時オプションなどによって chasenrc ファイルを指定することもできる。具体的には次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. 利用者のホームディレクトリにある .chasen2rc。
4. 利用者のホームディレクトリにある .chasenrc。
5. PREFIX/etc/chasenrc

設定項目一覧を以下に示す。このうち「DADIC」、「未知語品詞」、「品詞コスト」は必ず指定しなければならない。

1. 文法ファイルのディレクトリ

文法ファイル (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/ipadic/dic)
```

「文法ファイル」は省略することができ、その場合 chasenrc ファイルがあるディレクトリと同じディレクトリを指定したとみなされる。茶筌に付属の辞書 ipadic1.01 以降の chasenrc ファイルでは「文法ファイル」は省略されている。

2. システム辞書

ダブル配列辞書 (chadic.{da,lex,dat}) を、ファイル名から末尾の拡張子を除いたものを記述することによって指定する。複数組みの辞書を指定することもできる。また、相対パス、つまり “/” で始まらないパスを記述すると、文法ファイルと同じディレクトリを指定したとみなされる。例えば以下のように指定する。

```
(DADIC chadic
    /home/rikyu/mydic/chadic)
```

上の記述では、以下の二組の辞書ファイルが読み込まれる。

- (a) 文法ファイルと同じディレクトリにある chadic.{da,lex,dat}

(b) /home/rikyu/mydic/ にある chadic.{da,lex,dat}

辞書引きに際しては、これらの辞書の両方が用いられる¹。

Darts² によるダブル配列辞書を使うために「DADIC」を指定する。

(DADIC chadic)

上の記述では、文法ファイルと同じディレクトリにある chadic.da, chadic.lex, chadic.dat が読み込まれる。

使用する辞書の最大数は、32 個に設定されている。

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接続規則を適用するかを指示する。複数の品詞を指定した時は、それぞれの品詞について接続規則が適用される。

(未知語品詞 (名詞 サ変接続)) ; 1 個の品詞を指定
(未知語品詞 (名詞 サ変接続) (名詞 一般)) ; 複数の品詞を指定

4. 品詞のコスト

形態素解析プログラムでは、解析結果の優先情報をコストとして計算している。解析に曖昧性がある場合は、コストの総計が低いものを優先することになっている。「品詞コスト」では、各品詞のコストの倍率と「未知語」についてのコストを定義する。コストは正の整数値をとる。

(品詞コスト
((*) 1)
((未知語) 500)
((名詞) 2)
((名詞 固有名詞) 3)
)

同じ品詞に対してコストの定義が複数回指定されている場合は、後のものが優先される。上の記述では、「名詞」の形態素のコストは基本的には 2 倍になるが、「名詞-固有名詞」以下に細分類される名詞だけは形態素のコストが 3 倍になる。また、先頭の「(*)」の指定により、ここで明示的に定義されていない形態素のコストはすべて 1 倍 (そのままのコスト値) となる。未知語の形態素のコスト値はすべて 500 になる。

5. 接続コストと形態素コストの相対的な重みの定義

形態素解析におけるコストの計算は形態素のコストと接続のコストの総計として計算される。これら二種類のコストに異なる重みを掛けたい場合には、それを指定することができる。解析結果のコストはそれぞれのコストにここで指定された重みを乗じた値の総計として計算される。省略した場合の重みは 1 である。

(接続コスト重み 1) ; デフォルト値
(形態素コスト重み 1) ; デフォルト値

¹ 一組の辞書には同一の形態素の登録は行なわれないが、複数の辞書に同じ形態素が登録されている場合はあり得る。このような場合は、同じ形態素が複数得られることになる。

² <http://cl.aist-nara.ac.jp/%7etaku-ku/software/darts/>

6. コスト幅

形態素解析の過程において、常にコストが最低の結果を出すのではなく、ある程度のコスト幅を許容したい場合がある。この許容幅を指定することができる。コスト幅におさまるすべての解を出力するには `-m` オプションや `-p` オプションを使う。

(コスト幅 0) ; デフォルト値

コスト幅は `-w` オプションでも指定することができる。その場合、`-w` オプションで指定したものが優先される。

7. 未定義接続コストの定義

接続規則ファイルに接続規則が定義されていない形態素間の接続コストを指定する。未定義接続コストを設定しないか、あるいは 0 を指定すると、接続規則が定義されていない形態素どうしは決して接続しないという意味になる。デフォルトは 0。

(未定義接続コスト 500)

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 "%m\t%y\t%P-\n")

出力フォーマットは `-F` オプションでも指定することができる。その場合、`-F` オプションで指定したものが優先される。詳しくは 3 節を参照のこと。

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。“%S”を使うと解析文全体を表示できる。デフォルトは空文字列(何も表示しない)。

(BOS 文字列 "解析文: [%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。“%S”を使うと解析文全体を表示できる。デフォルトは“EOS\n”。

(EOS 文字列 "文末\n")

11. 空白品詞

茶釜は、半角の空白文字 (ASCII コード 32) とタブ (ASCII コード 9) を空白とみなし、これらを見捨てて解析する。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。例えば、以下のように設定すると、空白を「記号-空白」として出力する。

(空白品詞 (記号 空白))

なお、出力フォーマットを“%m”に設定して、空白品詞を指定する(品詞は何でもよい)と、解析文と全く同じ出力が得られることになる。

12. 注釈

ある文字列で始まりある文字列で終わる文字列を注釈のように扱い、その文字列の部分を無視して解析させることができる。解析結果には、その文字列が一つの形態素として出力される。

chasenrc ファイルには、開始文字列、終了文字列からなるリストと出力時の品詞名あるいはフォーマット文字列を記述する。終了文字列は省略することができ、その場合、開始文字列と一致する文字列自身を注釈として扱う。また、出力時の品詞名あるいはフォーマット文字列を省略するとその形態素についての情報を全く出力しなくなる。

```
(注釈 ((("<" ">") "%m\n")
  ("「” (記号 一般))
  ("」") (記号 一般))
  ("\"\" \"\\\"") (名詞 引用文字列))
  (" [ " " ]")
)
```

例えば、上のように記述すると、以下のように解析、出力される。

- のように “<” で始まり “>” で終わる文字列をそのまま出力する。
- “「” あるいは “」” を「記号-一般」として出力する。
- “hello(again)” のようにダブルクォーテーションで囲まれた文字列を「名詞-引用文字列」として出力する。
- “[ちゃせん]” のように “[” で始まり “]” で終わる文字列を無視して解析し、解析結果にはその文字列の情報を出力しない。

13. 連結品詞

ある品詞の形態素が連続して出現したときに、一つの形態素として連結して出力させるときに使用する。

```
(連結品詞 ((複合名詞) (名詞) (接頭詞 名詞接続) (接頭詞 数接続))
  ((記号)))
```

例えば、上の記述では以下のように品詞を連結する。

- 連続した「名詞」「接頭詞-名詞接続」「接頭詞-数接続」を連結し「複合名詞」として表示する。なお、「複合名詞」は品詞定義ファイル grammar.cha に記述しておく必要がある。
- 連続した「記号」を連結し「記号」として表示する。

14. 複合語出力

形態素辞書ファイル(.dic)内で定義した複合語について、複合語全体の形態素情報を出力する(“複合語”)か、複合語を構成する各単語の形態素情報を出力する(“構成語”)かを選択することができる。デフォルトは“複合語”。

```
(複合語出力 "複合語")
```

なお、複合語出力については -0c, -0s オプションによっても制御することができる。

15. 区切り文字

-j オプションをつけた時の文の区切り文字を並べ、一つの文字列にしたものを指定する(??節参照)。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。例えば

(区切り文字 "。、,!?.,!?")

と定義すると、全角文字の「。、,!? 」のいずれか、または半角文字の“.,!? ”(空白文字が入っていることに注意)のいずれかの文字が文の区切りとなる。

出力フォーマット

-F オプションや、chasenrc ファイルの「出力フォーマット」で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。

出力フォーマットの文字列の末尾に '\n' があれば、各形態素情報の表示ごとに改行を行い、文末の次に 'EOS' の1行を出力する。末尾に '\n' がなければ、1文中の形態素情報を1行で出力し、行末に改行を表示する。

また、出力フォーマットに '-f', '-e', '-c' を指定すると、それぞれ -f, -e, -c と同じ出力形式になる。

出力フォーマットの使用例をいくつかあげる。

- デフォルト (-f オプション) と同様の出力

"%m\t%y\t%M\t%U(%P-)\t%T_\t%F_\n" または "-f"

- 見出し、読み、品詞をタブで区切って表示

"%m\t%y\t%P-\n"

- 見出し語のみ

"%m\n"

- 分かち書き (見出し語を空白で区切って表示)

"%m_"

- 漢字かな変換

"%y"

- ルビつき表示。“漢字 (かな)” の形式で表示する。

"%r_()"

出力フォーマットの変換文字の一覧を以下に示す。

変換文字	機能
%m	見出し (出現形)
%M	見出し (基本形)
%y, %y1	読みの第一候補 (出現形)
%Y, %Y1	読みの第一候補 (基本形)
%y0	読み全体 (出現形)
%Y0	読み全体 (基本形)
%a	発音の第一候補 (出現形)
%A	発音の第一候補 (基本形)
%a0	発音全体 (出現形)
%A0	発音全体 (基本形)
%rABC	ルビつきの見出し (“A 漢字 B かな C” と表示)(1)
%i, %i1	付加情報の第一候補
%i0	付加情報全体
%Ic	付加情報 (空文字列か “NIL” なら文字c)(1)
%Pc	各階層の品詞を文字c で区切った文字列
%Pnc	1 ~ n(n:1 ~ 9) 階層目までの品詞を文字c で区切った文字列
%h	品詞の番号
%H	品詞文字列
%Hn	n(n:1 ~ 9) 階層目の品詞 (なければ最も深い階層)
%b	0(旧版との互換性のみ)
%BB	品詞細分類 (なければ品詞)
%Bc	品詞細分類 (なければ文字c)(1)
%t	活用型の番号
%Tc	活用型 (なければ文字c)(1)
%f	活用形の番号
%Fc	活用形 (なければ文字c)(1)
%c	形態素のコスト
%S	解析文全体
%pb	最適パスであれば “*”, そうでなければ “_”
%pi	パスの番号
%ps	パスの形態素の開始位置
%pe	パスの形態素の終了位置 +1
%pc	パスのコスト
%ppiC	前に接続するパスの番号を文字C で区切り列挙
%ppcC	前に接続するパスのコストを文字C で区切り列挙
%(B/STR1/STR2/	品詞細分類があればSTR1, なければSTR2(2)
%(I/STR1/STR2/	付加情報が “NIL” でも “”(空文字列) でもなければSTR1, そうでなければSTR2(2)
%(T/STR1/STR2/	活用があればSTR1, なければSTR2(2)
%(F/STR1/STR2/	%(T/STR1/STR2/ と同じ
%(U/STR1/STR2/	未知語ならSTR1\, そうでなければSTR2(2)
%(U/STR/	未知語なら”未知語”, そうでなければSTR(%U/未知語/STR/と同じ)(2)
%%	% そのもの

変換文字	機能
.	フィールド幅の指定
-	フィールド幅の指定
1-9	フィールド幅の指定
\n	改行文字
\t	タブ
\\	\ そのもの
\'	' そのもの
\"	" そのもの

1 ipadic では「行く (いく/ゆく)」のように形態素が複数の読みを持つ場合、その読みを「{イ/ユ}ク」のように、半角のブレースとスラッシュを使って表している。通常の読みの出力 (出力フォーマットの %y) では、その第一候補である「イク」が出力され、%y0 を使うと読み全体である「{イ/ユ}ク」が出力される。

1 A,B,C,c が空白文字の時は何も表示しない。

2 ‘/’ には任意の文字が使える。また、括弧 “(){}[]<>” を用いることもできる。以下に例をあげる。

- %?T#STR1#STR2#
- %?B(STR1)(STR2)
- %?U{STR1}/STR2/
- %U[STR]

4 形態素エントリの追加

4.1 各ファイルの編集

ipadic-X.X.X.tar.gz もしくは ipadic-sjis-X.X.X.zip をダウンロードし解凍する。
ダウンロードするファイルは以下にある。

- <http://chasen.aist-nara.ac.jp/stable/ipadic/>
- <http://chasen.aist-nara.ac.jp/stable/ipadic/win/>

前述したフォーマットにのっとり編集する。

- *.dic
形態素辞書
- connect.cha
接続表
- grammar.cha
品詞定義ファイル
- ctypes.cha
活用型定義ファイル
- cforms.cha
活用形定義ファイル

4.2 システム辞書の再コンパイル (UNIX 版)

品詞体系の変更など、システムの形態素辞書を編集した場合には、辞書の再コンパイルを必要とする。

1. `./configure` を実行する。

インストール先を手動で変更したい場合は以下のように実行する。

```
% ./configure --with-dicdir=/home/masayu-a
```

2. `make` を実行する。

```
% make
```

OS 標準の `make` を使うとコンパイルに失敗することがある。その場合は GNU `make` を使用する。

3. `root` 権限で `'make install'` を実行する。

```
# make install
```

デフォルトでは `/usr/local/share/chasen/dic/ipadic` にインストールされる (環境によって多少異なることもある)。

但し、ユーザーのディレクトリにインストールする場合には、`root` 権限は必要としない。

4.3 ユーザー辞書の再コンパイル (UNIX 版)

品詞体系の変更などを伴わない簡単な語彙の追加の場合には、ユーザー辞書を利用することができる。

まずユーザー辞書専用のディレクトリを用意する。

拡張子が `.dic` のファイル名をもつ形態素辞書ファイルを追加した後、次のようにコマンドを入力する。

```
% mkdir ~/mydic
% cd ~/mydic
% emacs Noun2.dic
(形態素情報を記述)
$ 'chasen-config --mkchadic'/makeda -i e chadic *.dic
```

`makeda` に付与される `i` オプションは文字コードを意味する。次の4つの文字コードが指定できる (`e`:EUC-JP, `s`:Shift_JIS, `w`:UTF-8, `a`:ISO-8859-1)。

```
% chasen-config --mkchadic
```

次に、`chasenrc` ファイルを `.chasenrc` というファイル名でホームディレクトリにコピーする。

```
% cd
% cp /usr/local/share/chasen/dic/ipadic/chasenrc .chasenrc
```

.chasenrc を編集し、以下のように「文法ファイル」を指定し、「DADIC」にユーザ辞書を追加する。

```
(文法ファイル /usr/local/share/chasen/dic/ipadic)
(DADIC chadic
    /home/masayu-a/mydic/chadic)
```

4.4 辞書の再コンパイル (Windows 版)

展開されたファイルを、Windows 版の茶筌がインストールされているディレクトリの中にある dic というディレクトリの下にコピーする。

単語を記述し終わったら、コマンドプロンプトで Makefile.bat を実行する。

```
C:\Program Files\chasen21> Makefile.bat
```

5 IPA 品詞体系

説明の書式

品詞名

以後、品詞名のことを「タグ」と呼ぶことがある。それぞれの品詞の説明の際に、以下の記号により注釈が付けられている。

品詞の解説

例：単語例

* 品詞の解説についての備考

& 読み、活用形についての備考

品詞名に関する注意事項

本日本語辞書は、IPA 品詞体系 (THiMCO97) に基づいているが、茶筌の辞書として組み込む際にいくつかの変更を行なった。品詞体系の特徴と変更点について以下にまとめる。

- 品詞は、多段の階層に分類されている。例えば「名詞 固有名詞 人名 姓」は、四段の階層よりなる品詞名である。以下では、これを「名詞-固有名詞-人名-姓」のようにハイフンで区切って表示する。茶筌 2.0 では、任意の段数の品詞階層の定義が可能になったので、これを直接文法ファイル (grammar.cha) に定義することができる。
- THiMCO97 では「動詞 一段 連用形 自立」のように、品詞の分類と活用型、活用形が混ざり合った形で定義されていた。茶筌では、品詞の分類の定義と活用に関する定義が分離されているので、これを「動詞-自立 一段 連用形」のように 3 つの項目 (品詞名、活用型、活用形) に分けて記述することにした。
- 品詞名の定義に用いられる分類名を以下の基準に従って変更した。

1. 「(助動詞語幹)」「(形容動詞語幹)」のように丸括弧を伴う名称の丸括弧をすべて除去した。
 2. 「動詞 接尾 (助動詞)」「形容詞 接尾 (助動詞)」として定義される「(助動詞)」の部分は冗長であるので、省略し「動詞-接尾」「形容詞-接尾」とした。
 3. 動詞の分類には「動詞」「動詞 非自立」「動詞 接尾」に大別されるが、茶筌の品詞階層の定義では、「動詞」という記述はすべての動詞を表すので、区別のため「動詞-自立」「動詞-非自立」「動詞-接尾」のように「自立」という細分類を追加した。
同様に、活用語以外の単語のための品詞名については、「名詞」「名詞 固有名詞」「名詞 固有名詞 人名」「名詞 固有名詞 人名 姓」のような分類を、それぞれ「名詞-一般」「名詞-固有名詞-一般」「名詞-固有名詞-人名-一般」「名詞-固有名詞-人名-姓」のように、「一般」という細分類を追加して、排他的に品詞の定義を行なった。
 4. 用言の活用形については「未然ナイ接続」「未然レル接続」「未然ウ接続」「連用タ接続」「連用マス接続」「連用タイ接続」…のように、後続する助動詞類に応じて細かく定義されていたが、個々の活用型については「未然」「連用」等で異なる語尾形を持つものは少ない。よって、活用形の名称は「未然形」「連用形」「基本形」「仮定形」「命令」を基本的な活用形とし、例外的な形のものに対してのみ、THiMCO97の活用形名を使用した。なお、茶筌では辞書出現形に対して「基本形」という活用形を与える仕様になっているため、THiMCO97の「見出し形」という活用形名を「基本形」という名前に変更した。
 5. 「未然ウ接続」は、五段活用動詞については助動詞「う」が接続し、その他の活用型の動詞については「よう」が接続するための活用であるとされていた。ここでは「う」のみを単語(助動詞)として認め「来よ(う)」「食べよ(う)」のように「よ」までを活用語尾とした。
- ipadic2.0以降の辞書では、単語に新たに「発音」フィールドが追加された。これは「日本語ディクテーション基本ソフトウェアの開発」グループの努力により添付されたものである。例えば、係助詞の「は」の読みは「ワ」「常識」の読みは「ジョーシキ」のように長音は「ー」によって示されている。また、綴りも品詞も等しいが読みだけが異なる単語、例えば「私(ワタシ/ワタクシ)」については、{ワタシ/ワタクシ}のようにすべての可能な読みを付与し、一つの語として登録した。

5.1 名詞

5.1.1 名詞-一般

普通名詞、あるいは、下位分類が未定の名詞。

5.1.2 名詞-固有名詞-一般

一般的な固有名詞、あるいは、下位分類が未定の固有名詞。

5.1.3 名詞-固有名詞-人名-一般

姓と名に分けられないもの、外国人名。あるいは、姓・名の決定が未定の人名。

例：「お市の方」

5.1.4 名詞-固有名詞-人名-姓

主に日本人の姓。

例：「山田」...

5.1.5 名詞-固有名詞-人名-名

主に日本人の名。

例：「太郎」...

5.1.6 名詞-固有名詞-組織

組織を表わす名称。

例：「通産省」「NHK」...

5.1.7 名詞-固有名詞-地域-一般

国名以外の地名を表わすもの。

例：「アジア」「バルセロナ」「京都」

5.1.8 名詞-固有名詞-地域-国

国の名前。

例：「日本」「オーストラリア」...

5.1.9 名詞-代名詞-一般

いわゆる代名詞，不定語。

例：「それ」「ここ」「あいつ」「あなた」「あちこち」「いくつ」「どこか」「なに」「みなさん」「みんな」「わたくし」「われわれ」...

5.1.10 名詞-代名詞-縮約

代名詞と係助詞「は」の組み合わせで，短縮した形<口語>。

例：「ありゃ」「こりゃ」「こりゃあ」「そりゃ」「そりゃあ」

5.1.11 名詞-副詞可能

曜日，月など時間を表す副詞的な用法を持つ名詞。量や割合などを表し副詞的に使うことのできる名詞。

例：「金曜」「一月」「午後」「少量」...

* 元の IPA 品詞体系では，「名詞-副詞可能」のうち実際に副詞的に使われてるものを「名詞 副詞可能 副詞的」，副詞的な使用が可能であるが，副詞的に用いられていないものを「名詞 副詞可能」とラベル付けされることになっているが，ここでは，文内の用法に関係なく副詞的に働き得るものをすべて「名詞-副詞可能」と呼ぶ。

5.1.12 名詞-サ変接続

格要素をとり、後ろに「する」「できる」「なさる」「くださる」などが後接することができるもの。

例：「インプット」「愛着」「悪化」「悪戦苦闘」「一安心」「下取り」...

* 「オノマトペ(+スル)」は、[副詞-助詞類接続]とした。

* 「名詞-一般」「名詞-サ変接続」の多重の用法があると考えられる時は、後者を優先して品詞名を付ける。

5.1.13 名詞-形容動詞語幹

いわゆる形容動詞語幹で「な」の前に現れるもの。

例：「健康」「安易」「駄目」「だめ」...

* 元の IPA 品詞体系では「名詞(形容動詞語幹)」となっていたが、第2階層の「(形容動詞語幹)」の括弧を取り除いた。

* 「名詞-一般」「名詞-形容動詞語幹」の多重の用法があると考えられる時は、後者を優先して品詞名を付ける。ただし、natureの意味の「自然」と「自然な」の用法で使われる語は、意味が違う(前者の連体用法は「自然の」)ため、前者を「名詞-一般」、後者を「名詞-形容動詞語幹」として別に登録した。

5.1.14 名詞-ナイ形容詞語幹

助動詞の「ない」の直前に現れて形容詞的な働きをする語

例：「申し訳」「仕方」「とんでも」「違い」...

* 元の IPA 品詞体系では形容詞とみなされていたが「申し訳-ない」「申し訳-ありません」「申し訳-ございません」のように派生するので、語幹として統一的に扱うことにした。ただし「ナイ形容詞語幹」として分類された語がすべてこのような用法を持つわけではない。

5.1.15 名詞-数

漢数字、算用数字、および「何(回)」「数(%)」「幾(夜)」。

例：「0」「1」「2」「何」「数」「幾」

5.1.16 名詞-非自立-一般

連体詞「の(格助詞)」、活用語の基本形に接続して使われるもののうち、以下の下位分類にあてはまらないもの。いわゆる形式名詞を含む。

* 普通名詞としての用法であれば、連体修飾を受けていても[非自立]ではない。

例：「あかつき」「暁」「かい」「甲斐」「気」「きらい」「嫌い」「くせ」「癖」「こと」「事」「ごと」「毎」「しだい」「次第」「順」「せい」「所為」「ついで」「序で」「つもり」「積もり」「点」「どころ」「の」「はず」「筈」「はずみ」「弾み」「拍子」「ふう」「ふり」「振り」「ほう」「方」「旨」「もの」「物」「者」「ゆえ」「故」「ゆえん」「所以」「わけ」「訳」「わり」「割り」「割」「ん<口語>」「もん<口語>」...

5.1.17 名詞-非自立-副詞可能

連体詞「の(格助詞)」, 活用語の基本形に接続して使われるもののうち, 副詞的に働くことが可能なもの.

* 文脈上で, 実際に副詞的に働いている場合には, IPA 品詞体系ではうしろに [副詞的] を付加することになっているが, それは省略した.

例: 「あいだ」「間」「あげく」「挙げ句」「あと」「後」「余り」「以外」「以降」「以後」「以上」「以前」「一方」「うえ」「上」「うち」「内」「おり」「折り」「かぎり」「限り」「きり」「つきり」「結果」「ころ」「頃」「さい」「際」「最中」「さなか」「最中」「じたい」「自体」「たび」「度」「ため」「為」「つど」「都度」「とおり」「通り」「とき」「時」「ところ」「所」「とたん」「途端」「なか」「中」「のち」「後」「ばあい」「場合」「日」「ぶん」「分」「ほか」「他」「まえ」「前」「まま」「儘」「俚」「みぎり」「矢先」...

5.1.18 名詞-非自立-助動詞語幹

連体詞「の(格助詞)」, 活用語の基本形に接続して使われるもののうち, 学校文法で助動詞とされる「ようだ」の語幹部分.

例: 「よう」「やう」「様(よう)」

* 元の IPA 体系では「名詞-非自立-(助動詞語幹)」と書かれていた.

5.1.19 名詞-非自立-形容動詞語幹

連体詞「の(格助詞)」, 活用語の基本形に接続して使われるもののうち「な(助動詞「だ」)の体言接続」と接続可能なもの.

例: 「みたい」「ふう」

* 元の IPA 体系では「名詞 非自立 (形容動詞語幹)」と書かれていた.

5.1.20 名詞-特殊-助動詞語幹

基本形に接続するもので, 学校文法で助動詞とされている「そうだ(伝聞)」の語幹部分.

例: 「そう」

* 元の IPA 体系では「名詞 特殊 (助動詞語幹)」と書かれていた.

5.1.21 名詞-接尾-一般

名詞,あるいは他の品詞の語幹[ガル接続]や[連用タイ接続]に接続して複合名詞を形成する語のうち, 下位の分類にあてはまらないもの. 一般に「接尾語」というよりも範囲が広く, 複合名詞の後ろ要素として用いられることが普通なもの.

例: 「おき」「かた」「方」「甲斐(がい)」「がかり」「ぎみ」「気味」「ぐるみ」「(~した)さ」「次第」「済(ず)み」「よう」「(でき)っこ」「感」「観」「性」「学」「類」「面」「用」...

5.1.22 名詞-接尾-人名

名詞のうち特に人名に接続して名詞を形成する接尾語。

例：「君」「様」「著」など。

5.1.23 名詞-接尾-地域

名詞のうち特に地域名に接続して名詞を形成する接尾語。

例：「町」「市」「県」など。

5.1.24 名詞-接尾-サ変接続

名詞に接続して名詞を形成する接尾語のうち「スル」に前接し得るもの。

例：「化」「視」「分け」「入り」「落ち」「買い」

5.1.25 名詞-接尾-助動詞語幹

他の品詞の連用形に接続し，学校文法で助動詞の語幹とされている「そうだ(様態)」の語幹部分。

例：「そう」

* 元の IPA 体系では「名詞 接尾 (助動詞語幹)」と書かれていた。

5.1.26 名詞-接尾-形容動詞語幹

他の名詞や活用語の連用形に接続する接尾語で「だ」([助動詞特殊・ダ])に前接するもの。

例：「的」「げ」「がち」

* 元の IPA 体系では「名詞 接尾 (形容動詞語幹)」と書かれていた。

5.1.27 名詞-接尾-副詞可能

他の名詞に接続する接尾語で，副詞的に働くことが可能なもの。

* IPA 品詞体系では，文脈上，実際に副詞的に働いているものは，うしろに[副詞的]と付加してあるが，ここでは具体的な用法にかかわらず，副詞的な使用が可能なものをすべてこの分類とした。

例：「後(ご)」「以後」「以降」「以前」「前後」「中」「末」「上」「時(じ)」

5.1.28 名詞-接尾-助数詞

数に接続して名詞を形成する接尾。一般の「助数詞」よりも範囲が広く，数に接続する普通名詞も含まれる。

例：「個」「つ」「本」「冊」「パーセント」「cm」「kg」「カ月」「か国」「区画」「時間」「時半」...

* IPA 品詞体系では，これらのうち副詞的に用いられているものに「名詞 接尾 助数詞 副詞的」というタグを与えているが，これは用法に関するタグであるため，本体系では含めなかった。

5.1.29 名詞-接尾-特殊

主に用言につく特殊な接尾辞として新たに定義した。

例：「(楽し)さ」「(考え)方」

* IPA 品詞体系では、「名詞 接尾」に分類されていた。

5.1.30 名詞-接続詞的

単語と単語を接続する接続詞的な働きをするもの。

例：「(日本)対(アメリカ)」「対(アメリカ)」「(3)対(5)」「(女優)兼(主婦)」

5.1.31 名詞-動詞非自立的

[助詞-接続助詞]の「て」に接続するもので、意味的には動詞的なもの。

例：「ごらん」「ご覧」「御覧」「頂戴」

注 IPA 品詞体系には、単語への分割が不可能なもの、および、ことわざ、漢詩、方言、英語などを表すタグとして「名詞 引用文字列」が用意されている。また、数式を表すためのタグ「名詞 数式」が用意されている。これらは品詞タグとは考えにくいいため、本体系では正式な品詞名としては採用しない立場を取るが、現在は「名詞 引用文字列」として「いわく」だけが残されている。

5.2 接頭詞

5.2.1 接頭詞-名詞接続

名詞(形容動詞語幹を含む)に前接する接頭語のうち、数に接続するもの以外。

例：「お(水)」「某(氏)」「同(社)」「故(～氏)」「高(品質)」「お(見事)」「ご(立派)」

5.2.2 接頭詞-数接続

名詞に前接する接頭語のうち、数に接続するもの。

例：「約」「およそ」「毎時」など

5.2.3 接頭詞-動詞接続

動詞の命令形あるいは[動詞 連用形]+「なる/なさる/くださる」に前接する接頭語。

例：「お(読みなさい)」「お(座り)」

5.2.4 接頭詞-形容詞接続

形容詞に前接する接頭語。

例：「お(寒いですねえ)」「バカ(でかい)」

5.3 動詞

活用形に関する注意

未然形 THiMCO97 では以下のように細かく分類されているが、語尾形に変化のない限り、「未然形」に統一した。

- 未然レル接続
- (ラ)レル, - (サ)セルに接続するもの。
例: 「読ま」「さ」...
- 未然ナイ接続
- ナイに接続するもの。
例: 「読ま」「し」...
- 未然又接続
- 又, - (サ)シメルに接続するもの。
例: 「読ま」「せ」「来」...
- 未然ウ接続
- (ヨ)ウに接続するもの。
例: 「読も」「し」...
& ipadic1.0 以降では、助動詞ウが接続するものと定義した。例えば、「する」については、「しよ」が未然ウ接続。

連用形 例外的な語尾以外はすべて「連用形」という名称に統一した。

- 連用マス接続
- マスに接続するもの。
例: 「読み」「し」「なさい」...
- 連用タイ接続
- タイ, - ソウ, - ツライ, - 方(かた), 読点などに接続するもの。
例: 「読み」「し」「なさり」「向かひ」「習ひ」...
- 連用夕接続
- タ, - テに接続するもの。
例: 「読ん」「書い」「行っ」「問う」...

基本形 THiMCO97 では、「見出し形」と呼ばれているもの。

- # 句点, 体言, - マイなどに接続するもの。
例: 「読む」「なさる」「問う」...

仮定形 THiMCO97 では、「仮定バ接続」と呼ばれている。

- # - バ, - ドモに接続するもの。
例: 「読め」「すれ」...

命令 i # 力変・五段ラ行特殊の命令形。およびサ変・スルの命令形「せよ」の口語形。

例：「来い」「なさい」「せい」...

命令 e # 五段の命令形，文語已然形，一段動詞の語幹止め命令用法（「くれ」のみ）。

例：「読め」「(とは)いえ」「(程度の差こそ)あれ」「(やめて)くれ」...

* 「(やめて)くれ」は「(やめて)くれろ」の「ろ」が落ちた形！「くれる」は一段動詞の中の特殊活用型とすべきものである。なお「(やめて)(お)くれ(なさい)」の「くれ」は[動詞-非自立 一段 連用タイ接続]であり，別のものであり，この口語形は「おくんない」となる。

命令 y o # 一段・サ変・文語(力変)の命令形で「y o」で終わるもの。

例：「せよ」「みよ」「来よ」...

命令 r o # 一段・サ変の命令形で「r o」で終わるもの。

例：「しろ」「みろ」...

ベキ接続 # 「ベキ」につづく形，サ変の場合のみ。

例：「す」...

仮定縮約 1 # 仮定バ接続と「バ」とが合わさって短縮した形<口語>。

例：「分かれりゃ」

体言接続 # 文語の場合のみ。見出し形と異なる形があるもの。

例：「助くる」(cf.「助く」)

体言接続特殊 # 「る」で終る動詞が「の」などに接続する場合に音便化した形<口語>。

例：「(何)すん(の?)」

体言接続特殊 2 # 「来る」「する」「とる」などの「体言接続特殊」の語末の「ん」が欠落したもの。<口語>。

動詞の活用型一覧(現代語)

【活用形】の位置には次のものが入る(【】の記号はない)

5.3.1 動詞-自立 力変【活用形】

例：「くる」「来る」「やってくる」「やって来る」

5.3.2 動詞-非自立 力変【活用形】

例：「(て)くる」「(て)来る」

5.3.3 動詞-自立 サ変・スル【活用形】

「する」，および[名詞 サ変接続]に接続する「する」。

例：「する」

5.3.4 動詞-自立 サ変・ - スル【活用形】

和語系のサ変動詞 .

例：「接する」...

* 「 - し + ない」「 - せ + られる」「 - せ + ぬ」「 - し + よう」「 - する」「 - すれ + ば」「 - せよ」「 - しろ」の形だけを[動詞 サ変・スル]とした。「 - し + 」 「 - し + た」「 - し + たい」などの連用形はすべて[動詞 五段・サ行]とした .

5.3.5 動詞-自立 サ変・ - ズル【活用形】

和語系のザ変動詞 .

例：「信ずる」...

* 「 - ぜ + られる」「 - ぜ + ぬ」「 - ずる」「 - ずれ + ば」「 - ぜよ」「 - ず + べし」の形だけを[動詞 サ変・ズル]とした。「 - じ + ない」「 - じ + よう」の未然形および「 - じ + 」 「 - じ + た」「 - じ + たい」などの連用形 , および「 - じろ」の命令形は[動詞 一段]とした .

5.3.6 動詞-自立 一段【活用形】

いわゆる上一段活用および下一段活用 .

例：「着る」

* 「病める」は[基本形]のみ .

5.3.7 動詞-非自立 一段【活用形】

例：「あげる」「うる」「える」「得る」「おえる」「終える」「おおせる」「かねる」「兼ねる」「かける」「きる」「切れる」「すぎる」「過ぎる」「そこねる」「損ねる」「そびれる」「そめる」「初める」「つける」「つづける」「続ける」「(お読み)できる」「(お読み)出来る」「はじめる」「始める」「(て)いる」「(～しては)いけ(ない)」「(て)くれる」「(て)差し上げる」「(て)のける」「(て)みる」「(て)みせる」「(て)もらえる」「(て)る<口語>」

* 「(～しては)いけ(ない)」の基本形は「いける」 .

* 「(勉強)できる」は[非自立]としない .

* 「うる」は「うる」([基本形])と「うれ」([仮定形])のみ。「う」は[動詞 文語 基本形]とする .

5.3.8 動詞-接尾 一段【活用形】

学校文法では助動詞と呼ばれているもの .

例：「させる」「せる」「しめる」「しむる」「られる」「れる」

5.3.9 動詞-自立 五段・カ行イ音便【活用形】

五段カ行で[助詞 接続助詞]の「て」に接続するときイ音便になるもの .

例：「解く」「聞く」...

5.3.10 動詞-非自立 五段・カ行イ音便【活用形】

例：「つづく」「続く」「ぬく」「抜く」「(て)いただく」「(て)頂く」「(て)おく」「とく<口語>」「どく<口語>」

5.3.11 動詞-非自立 五段・カ行促音便【活用形】

五段カ行で[助詞 接続助詞]の「て」に接続するときに促音便になるもの。

例：「いく」「行く」「ゆく」

* 「ゆく」には「ゆっ(て)」の形はないが、このタグを振る「ゆき(て)」は[動詞 文語 連用タ接続]とする。

5.3.12 動詞-非自立 五段・カ行促音便【活用形】

例：「いく」「行く」「ゆく」「く<口語>」

* 「ゆく」には「ゆっ(て)」の形はないが、このタグを振る「ゆき(て)」は[動詞 文語 連用タ接続]とする。

5.3.13 動詞-自立 五段・ガ行【活用形】

五段ガ行で[助詞 接続助詞]の「て」に接続するときにイ音便になるもの。

例：「継ぐ」「急ぐ」...

5.3.14 動詞-自立 五段・サ行【活用形】

五段サ行で[助詞 接続助詞]の「て」に接続するときに音便化しないもの。

例：「話す」...

5.3.15 動詞-非自立 五段・サ行【活用形】

例：「いたす」「致す」「だす」「出す」「つくす」「尽くす」「直す」

5.3.16 動詞-自立 五段・タ行【活用形】

五段タ行で[助詞 接続助詞]の「て」に接続するときに促音便になるもの。

例：「持つ」...

5.3.17 動詞-自立 五段・ナ行【活用形】

五段ナ行で[助詞 接続助詞]の「て」に接続するときにハツ音便になるもの。

例：「死ぬ」

5.3.18 動詞-自立 五段・バ行【活用形】

五段バ行で [助詞 接続助詞] の「て」に接続するときにハツ音便になるもの。

例：「呼ぶ」...

5.3.19 動詞-自立 五段・マ行【活用形】

五段マ行で [助詞 接続助詞] の「て」に接続するときにハツ音便になるもの。

例：「進む」...

5.3.20 動詞-非自立 五段・マ行【活用形】

例：「こむ」「込む」

5.3.21 動詞-自立 五段・ラ行【活用形】

五段ラ行で [助詞 接続助詞] の「て」に接続するときに促音便になるもの。

例：「切る」「なる」...

5.3.22 動詞-非自立 五段・ラ行【活用形】

例：「おわる」「終る」「終わる」「かかる」「きる」「切る」「しぶる」「渋る」「まいる」「まわる」「回る」「やがる」「(せねば/しては)ならない」「(て)ある」「(て)おる」「(て)まわる」「(て)回る」「(て)やる」「ちやる<口語>」「じゃる<口語>」「ぢやる<口語>」

* 「ならない」の基本形は「なる」

5.3.23 動詞-接尾 五段・ラ行【活用形】

例：「がる」

5.3.24 動詞-自立 五段・ラ行特殊【活用形】

五段ラ行で、助動詞「ます」に接続する形および命令形が「-い」の形になるもの。

例：「いらっしゃる」「おっしゃる」「仰言る」「くださる」「下さる」「なさる」「ござる」

5.3.25 動詞-非自立 五段・ラ行特殊【活用形】

例：「(お読み)なさる」「(お読み)くださる」「(お読み)下さる」「(て)くださる」「(て)下さる」「(て)いらっしゃる」「(て)らっしゃる<口語>」

5.3.26 動詞-自立 五段・ワ行ウ音便【活用形】

五段ワ行で [助詞 接続助詞] の「て」に接続するときにウ音便になるもの。

例：「問う」「乞う」「沿う(て)」「ゆう(て)」「食う(て)」「すう(て)」「負う(て)」

- * [動詞 五段・ワ行促音便]の動詞については、「て」に接続する活用語尾が「う」になっている場合のみ、このタグを振り、それ以外は[動詞 五段・ワ行促音便]を振る(人手修正データ中に現れたものは「ゆう」「食う」「すう」「負う」)

5.3.27 動詞-非自立 五段・ワ行ウ音便【活用形】

例：「たまう」「給う」

5.3.28 動詞-自立 五段・ワ行促音便【活用形】

五段ワ行で[助詞 接続助詞]の「て」に接続するときに促音便になるもの。

例：「言う」「ゆう」「食う」「負う」「憂う」

- * 「憂う」には「憂って」はないが、このタグを振る(人手修正データ中に現れたものは「憂い(、)」の形のみ)
- * [動詞 五段・ワ行促音便]の動詞については、「て」に接続する活用語尾が「う」になっている場合のみ、[動詞 五段・ワ行ウ音便]を振る。

5.3.29 動詞-非自立 五段・ワ行促音便【活用形】

例：「あう」「合う」「そこなう」「損なう」「(て)しまう」「(て)もらう」「じゃう<口語>」「じまう<口語>」「ちまう<口語>」「ちゃう<口語>」

動詞の活用型一覧(文語)

IPA 品詞体系では、文語の活用型の細分類は行われていない。IPADIC2.4以降では、四段、上二、下二などの各活用型を定義したが、辞書中の実例はまだ充実していない。活用体系が文語残存のものと、口語ではあるが歴史仮名づかいで示されているものを含む。

5.3.30 動詞-自立 四段・八行【活用形】

例：「いふ」「云ふ」「向かふ」「習ふ」「思ふ」「能ふ」など。

- * 四段にはその他「カ行」「ガ行」「サ行」「タ行」「バ行」「マ行」「ラ行」が定義されている。

5.3.31 動詞-自立 ラ変【活用形】

例：「あり」「なり」「しかり」

5.3.32 動詞-自立 上二・八行【活用形】

- * その他「ダ行」がある。

5.3.33 動詞-自立 下二・ア行【活用形】

* その他「カ行」「ガ行」「サ行」「ザ行」「タ行」「ダ行」「ナ行」「ハ行」「バ行」「マ行」「ヤ行」「ラ行」「ワ行」「得」がある。

5.3.34 動詞-自立 一段・得ル【活用形】

古語「得る」の活用型。基本形と仮定形のみ。

5.4 形容詞

「見出し形」「仮定バ接続」「文語見出し形」をそれぞれ「基本形」「仮定形」「文語基本形」と呼ぶ以外は、ほぼ THiMCO97 で用いられている活用形名を用いた。なお、形容詞の活用型を「形容詞・アウオ段」「形容詞・イ段」「形容詞・文語」に分類した。

未然又接続

- 又に接続するもの。

例：「寒から」...

未然ウ接続

- ウに接続するもの。

例：「寒かる」...

連用夕接続

- 夕に接続するもの。

例：「寒かつ」...

連用テ接続

- テ、- ナイ、- ナル、- スル、読点に接続するもの。

例：「寒く」...

連用ゴザイ接続

- ゴザイマスに接続するもの。

例：「寒う」「大きゅう」「のう」...

基本形

句点、体言などに接続するもの。

例：「寒い」「大きい」「ない」...

体言接続

文語活用で体言に接続するもの .

例：「寒き」「なき」...

& 基本形には「 - イ」の形を入れた .

仮定形

- バに接続するもの .

例：「寒けれ」「なけれ」...

& THiMCO97 では「仮定バ接続」と呼ばれていた .

命令

文語活用で命令形のもの .

例：「よかれ」「美しかれ」...

& 終止形には「 - イ」の形を入れた .

文語基本形

- シで終わるもの .

例：「良し」「遠し」「やむなし」...

仮定縮約 1

仮定バ接続と「バ」とが合わさって短縮した形 1 <口語> .

例：「欲しけりゃ」「(それが) なけりゃ(困る)」

仮定縮約 2

仮定バ接続と「バ」とが合わさって短縮した形 2 <口語> .

例：「(それが) なきゃ(困る)」

ガル接続

- ガル , - ゲ , - ソウに接続するもの .

例：「寒」「悲し」...

【活用形】の位置には次のものが入る (【】の記号はない)

5.4.1 形容詞-自立 形容詞・アウオ段【活用形】

形容詞の活用型のうち、語幹の最後の母音がアウオのいずれかで終わるもの。

例：「青い」「赤い」「厚い」「暑い」「熱い」...

* IPA 品詞体系では、「ない」の文語見出し形「なし」を形容詞の文語型活用の見出し形として定義しているが、本体系では「形容詞・アウオ段(あるいは、形容詞・イ段)」型の「文語基本形」という活用形として定義している。同様に、IPA 体系で形容詞の文語型活用の体言接続と定義されている「悪しき」などは、本体系では他の形容詞と同様に「体言接続」として定義した。

5.4.2 形容詞-自立 形容詞・イ段【活用形】

形容詞の活用型のうち、語幹の最後の母音がイで終わるもの。

例：「哀しい」「楽しい」「頼もしい」...

5.4.3 形容詞-自立 形容詞・イイ【活用形】

例：「いい」「ええ」...

5.4.4 形容詞-自立 形容詞・不変化型【活用形】

形容詞の中で「基本形」の用法しかもたないもの。

例：「カッコいい」

5.4.5 形容詞-非自立 形容詞・アウオ段【活用形】

動詞の[連用タイ接続]あるいは[連用タ接続]に後接する形容詞。

例：「がたい」「難しい」「づらい」「にくい」「やすい」「(て)よい」「(て)良い」

5.4.6 形容詞-非自立 形容詞・イ段【活用形】

動詞の[連用タイ接続]あるいは[連用タ接続]に後接する形容詞。

例：「らしい」「(て)ほしい」「(て)欲しい」

5.4.7 形容詞-非自立 形容詞・イイ【活用形】

例：「いい」

5.4.8 形容詞-非自立 形容詞・不変化型【活用形】

動詞の[連用タイ接続]あるいは[連用タ接続]に後接する形容詞で「基本形」の用法しかもたないもの。

5.4.9 形容詞-接尾 形容詞・アウオ段【活用形】

学校文法では助動詞とされるもの。

例：「(食べ)たい」

5.4.10 形容詞-接尾 形容詞・イ段【活用形】

学校文法では助動詞とされるもの。

例：「(嫌味)たらしい」

5.5 副詞

5.5.1 副詞-一般

必ず後ろで切れるもの，連体修飾が不可能なもの。

例：「あいかわらず」「多分」など。

5.5.2 副詞-助詞類接続

「の」「は」「に」「な」「する」「だ」などが後続することが可能な副詞。

例：「こんなに」「そんなに」「あんなに」「なにか」「なんでも」

5.6 連体詞

5.6.1 連体詞

名詞を修飾する形しかもたないもの。

例：「この」「その」「あの」「どの」「いわゆる」「なんらかの」「何らかの」「いろんな」「こういう」「そういう」「ああいう」「どういう」「こんな」「そんな」「あんな」「どんな」「大きな」「小さな」「おかしな」「ほんの」「たいした」「(-も)さる(ことながら)」「微々たる」「堂々たる」「単なる」「いかなる」「我が」「同じ」「亡き」...

5.7 接続詞

5.7.1 接続詞

独立に現れる接続詞。

例：「が」「けれども」「そして」「じゃあ」「それどころか」...

5.8 助詞

5.8.1 助詞-格助詞-一般

いわゆる格助詞。

* 「にて」も格助詞に含めた「の」には格助詞としての用法と名詞と名詞を接続するいわゆる「AのB」用法があるが、後者は「助詞 連体化」とした。

例：「から」「が」「で」「と」「に」「へ」「より」「を」「の」「にて」

5.8.2 助詞-格助詞-引用

名詞、人物の台詞「」、会議の決定事項、理由、判決、推測表現等の直後の「と」。

例：「(～だ)と(述べた。)」(～である)と(して執行猶予...)

5.8.3 助詞-格助詞-連語

格助詞と動詞との連語で、主に格助詞に相当するような働きを持つもの。

例：「という」「といった」「とかいう」「として」「とともに」「と共に」「でもって」「にあたって」「に当たって」「に当って」「にあたり」「に当たり」「に当り」「に当たる」「にあたる」「において」「に於いて」「に於て」「における」「に於ける」「にかけ」「にかけて」「にかんし」「に関し」「にかんして」「に関して」「にかんする」「に関する」「に際し」「に際して」「にしたがい」「に従い」「に従う」「にしたがって」「に従って」「にたいし」「に対し」「にたいして」「に対して」「にたいする」「に対する」「について」「につき」「につけ」「につけて」「につれ」「につれて」「にとり」「にとり」「にまつわる」「によって」「に依って」「に因って」「により」「に依り」「に因り」「による」「に依る」「に因る」「にわたって」「にわたる」「をもって」「を以って」「を通じ」「を通じて」「を通して」「をめぐって」「をめぐり」「をめぐる」「って<口語>」「ちゅう<関西弁「という」>」「(何)ていう(人)<口語>」「っていう<口語>」「といふ」「とかいふ」

5.8.4 助詞-接続助詞

例：「から」「からには」「が」「けれど」「けれども」「けど」「し」「つつ」「て」「で」「と」「ところが」「どころか」「とも」「ども」「ながら」「なり」「ので」「のに」「ば」「ものの」「や(～した)」「やいなや」「(ころん)じゃ(いけない)<口語>」「(行っ)ちゃ(いけない)<口語>」「(言っ)たって(しかたがない)<口語>」「(それがなく)ったって(平気)<口語>」

5.8.5 助詞-係助詞

例：「こそ」「さえ」「しか」「すら」「は」「も」「ぞ」

5.8.6 助詞-副助詞

例：「がてら」「かも」「くらい」「位」「ぐらい」「しも」「(学校)じゃ(これが流行っている)<口語>」「(それ)じゃあ(よくない)<口語>」「ずつ」「(私)なぞ」「など」「(私)なり(に)」「(先生)なんか(大嫌い)<口語>」「(私)なんぞ」「(先生)なんて(大嫌い)<口語>」「のみ」「だけ」「(私)だって<口語>」「だに」「(彼)ったら<口語>」「(お茶)でも(いかが)」「等(とう)」「(今後)とも」「ばかり」「ばっか<口語>」「ばっかり<口語>」「ほど」「程」「まで」「迄」「(誰)も(が)〔助詞-格助詞〕および〔助詞-係助詞〕の前に位置する「も」)

5.8.7 助詞-並立助詞

例：「と」「たり」「だの」「だり」「とか」「なり」「や」「やら」

5.8.8 助詞-終助詞

例：「かい」「かしら」「さ」「ぜ」「(だ)っけ<口語>」「(とまってる)で<方言>」「な」「ナ」「なあ<口語>」「ぞ」「ね」「ネ」「ねえ<口語>」「ねえ<口語>」「ねん<方言>」「の」「のう<口語>」「や」「よ」「ヨ」「よお<口語>」「わ」「わい<口語>」

* 終助詞の「や」は「(まあいい)や」「(すごい)や」など。関西方言の断定の「や」は、不変化活用の助動詞として扱う。

5.8.9 助詞-副助詞 / 並立助詞 / 終助詞

「か」のうち、副助詞、並立助詞、終助詞いずれかわからないもの、たとえば、次の(a)(b)(c)のようなもの。

(a)「AかBか」型。例：「(国内で運用する)か(海外で運用する)か(。)」

(b) 副詞節中。例：「(幸いという)か(、死者はいなかった。)」 「(祈りが届いたせい)か(、試験に合格した。)」

(c)「かのように」。例：「(何もなかった)か(のように振る舞った。)」

例：「か」

* 最新のIPA品詞体系では、これをさらに「副助詞」「並立助詞」「終助詞」に細分類しているが、本体系では区別しなかった。

5.8.10 助詞-連体化

[名詞]に接続して体言にかかる「の」。

* THiMCO97では、この用法の「の」も格助詞に分類されている。

5.8.11 助詞-副詞化

擬音語、擬声語、擬態語及びそれに類する名詞や副詞の直後の「に」「と」。

例：「に」「と」

* ただし、「する」「なる」に係っているもので、状態変化を表す場合は格助詞とする。

5.8.12 助詞-特殊

以上の分類にあてはまらないもの。短歌や俳句等に用いられる助詞などを含む。

例：「かな」「けむ」「(～したろう)に」「(あんた)にや(わからん)」「(俺)ん(家)」

5.8.13 助詞-間投助詞

IPA 辞書には元々なかった品詞．文語的な用法をもつ．ipadic2.4 以降で定義した．

例：「(松島)や」

5.9 助動詞

5.9.1 助動詞 五段・ラ行アル【活用形】

動詞型の活用の助動詞「である」「ではある」などの「ある」．

例：「ある」

5.9.2 助動詞 五段・ラ行ゴザル【活用形】

動詞型の活用の助動詞「ござる」．

例：「ござる」

5.9.3 助動詞 形容詞・イ段【活用形】

形容詞型活用の助動詞．

例：「らしい」

5.9.4 助動詞 特殊・ナイ【活用形】

否定の助動詞「ない」の活用型．

例：「ない」

5.9.5 助動詞 特殊・タ【活用形】

完了を表す助動詞「た」の活用型．

例：「た」「だ」

* 「(学ん)だ」「(泳い)だ」のように五段活用のガ行，ナ行，バ行，マ行に接続する場合には表層形が「だ」になるため，本体系では，それぞれ別の語幹をもつ形態素として定義した．

5.9.6 助動詞 特殊・ダ【活用形】

断定の助動詞「だ」の活用型．

例：「だ」

5.9.7 助動詞 特殊・デス【活用形】

断定の助動詞「です」の活用型．

例：「です」

5.9.8 助動詞 特殊・ドス【活用形】

断定の助動詞「どす」の活用型 .

例：「どす」

5.9.9 助動詞 特殊・ジャ【活用形】

断定の助動詞「じゃ」の活用型 .

例：「じゃ」

* 断定の「だ」が鈍ったもの .

5.9.10 助動詞 特殊・マス【活用形】

謙譲・丁寧を表わす助動詞「ます」の活用型 .

例：「ます」

5.9.11 助動詞 特殊・ヌ【活用形】

否定の助動詞「ぬ」の活用型 .

例：「ぬ」

5.9.12 助動詞 特殊・ヤ【活用形】

関西方言の断定を表す助動詞 .

例：「(そう)や」

5.9.13 助動詞 不変化型【活用形】

現代では活用しない助動詞 . 活用が想定できない口語や方言も含む .

例：「う」「まい」「(いざ行か)ん(む)」「(去り)ぬ」「(わから)ん<口語>」「(賜ラ)ン」「~(美しい/学生)じゃん<口語>」「(いい)っす<口語>」「(負けてなら)じ」など .

5.9.14 助動詞 文語・??【活用形】

文語の助動詞 . 現在定義されている活用型は次の通り「文語・ベシ」「文語・ゴトシ」「文語・ナリ」「文語・マジ」「文語・シム」「文語・キ」「文語・ケリ」「文語・ル」「文語・リ」

例：「べし」「ごとし」「如し」「たり」「なり」「まじ」「き」「けり」「り」「る」

* IPA 品詞体系では「じ」のための活用型が用意されているが、実際には不変化であるので、ここでは不変化型とした .

5.10 感動詞

5.10.1 感動詞

感動詞 . あいさつなど .

例 : 「おはよう」「おはようございます」「こんにちは」「こんばんは」「ありがとう」「どうもありがとう」「ありがとうございます」「いただきます」「ごちそうさま」「さよなら」「さようなら」「はい」「いいえ」「ごめん」「ごめんなさい」...

5.11 記号

5.11.1 記号-一般

以下の分類以外の一般的な記号 .

例 : 「」 「」 「@」 「\$」 「〒」 「」 「+」 など .

5.11.2 記号-アルファベット

英語のアルファベット . 大文字 , 小文字 .

例 : 「A」「a」

5.11.3 記号-句点

いわゆる句点 .

例 : 「.」「。」

5.11.4 記号-読点

いわゆる読点 .

例 : 「,」「,」

5.11.5 記号-空白

全角の空白文字 (画面上には見えない) .

5.11.6 記号-括弧開

例 : 「(」 「{」 「『」 「“」 「『」 「【」 ...

5.11.7 記号-括弧閉

例 : 「)」 「}」 「』」 「”」 「』」 「】」 ...

5.12 フィラー

5.12.1 フィラー

話し言葉で起こるあいずちや挿入的な音声ことば

例：「あの」「うんと」「えと」

5.13 その他

5.13.1 その他-間投

[名詞-接尾]や[助詞-終助詞]としくいもの。

例：「(だ)ァ」

参考文献

- [1] 益岡隆志, 田窪行則: 『基礎日本語文法-改訂版-』くろしお出版, 1992.
- [2] 松本裕治, 黒橋禎夫, 宇津呂武仁, 妙木裕, 長尾真: 「日本語形態素解析システム JUMAN 使用説明書 version 2.0」, NAIST Technical Report, NAIST-IS-TR94025, 1994.
- [3] 山下達雄: 「規則と確率モデルの統合による形態素解析」, 奈良先端科学技術大学院大学修士論文, NAIST-IS-MT9551119, March 1997.
- [4] 山下達雄, 松本裕治: 「コスト最小法と確率モデルの統合による形態素解析」, 情報処理学会研究報告 96-NL-119, May 1997.
- [5] 北内 啓, 山下 達雄, 松本 裕治: 「日本語形態素解析システムへの可変長接続規則の実装」, 言語処理学会第三回年次大会論文集, pp.437-440, 1997.
- [6] 「研究開発用知的資源タグ付きテキストコーパス報告書」平成9年度, テキストサブワーキンググループ, 技術研究組合 新情報処理開発機構, 1998.
- [7] 松田 寛: 「品詞タグ付きコーパス作成支援環境の構築」, 奈良先端科学技術大学院大学修士論文, NAIST-IS-MT9851103, March 1999.
- [8] 北内 啓, 宇津呂 武仁, 松本 裕治: 「誤り駆動型の素性選択による日本語形態素解析の確率モデル学習」, 情報処理学会論文誌 Vol. 40, No. 5, p.p.2325-2337, May 1999.
- [9] 松田 寛, 桐山 和久, 山田 悟史, 吉野 圭一, 松本裕治: 「部分形態素解析を用いたコーパスの品詞体系変換」, 情報処理学会研究報告 99-NL-134, p.p.23-30, Nov. 1999.
- [10] Masayuki Asahara: Extended Statistical Model for Morphological Analysis, 奈良先端科学技術大学院大学修士論文, NAIST-IS-MT9851001, March 2000.
- [11] 松田 寛, 松本 裕治: 「品詞タグ付きコーパス作成支援 GUI ツール VisualMorphs」, 情報処理学会研究報告 2000-NL-137, p.98, June, 2000.
- [12] 浅原 正幸, 松本 裕治: 「統計的日本語形態素解析に対する拡張 HMM モデル」, 情報処理学会研究報告 2000-NL-137, p.p.39-46, June, 2000.

- [13] Masayuki Asahara, Yuji Matsumoto: Extended Models and Tools for High-performance Part-of-Speech Tagger, Proceedings of COLING 2000, July, 2000.
- [14] 浅原 正幸, 松本 裕治: 「誤り駆動による統計的品詞タグづけモデルの拡張」, 情報処理学会研究報告 2000-NL-139, p.p.25-32, Sep. 2000.
- [15] 松本 裕治: 「形態素解析システム『茶釜』」, 情報処理 Vol.41 No.11, p.p.1208-1214, Nov. 2000.
- [16] 伝 康晴, 浅原 正幸: 「リレーショナル・データベースによる統合的言語資源管理環境」, ワークショップ「話し言葉の科学と工学」, Feb. 2001.
- [17] 松本 裕治, 伝 康晴: 「話し言葉の形態素解析」, 情報処理学会研究報告 2001-NL-143, p.p.49-54, May, 2001.
- [18] Masayuki Asahara and Ryuichi Yoneda and Yuji Matsumoto: 「Use of a Relational Database in the Development and Maintenance of Linguistic Resources for Statistical Japanese Morphological Analysis」, IRCS Workshop on Linguistic Databases, Dec. 2001.
- [19] 浅原 正幸, 松本 裕治: 「形態素解析のための拡張統計モデル」, 情報処理学会論文誌, Vol.43 No.03, Mar. 2002.
- [20] 浅原 正幸, 米田 隆一, 山下 亜希子, 伝 康晴, 松本 裕治: 「リレーショナルデータベースによる品詞タグつきコーパスの管理手法」, SIG-SLUD-34 Mar. 2002.
- [21] Masayuki Asahara and Ryuichi Yoneda and Akiko Yamashita and Yasuharu Den and Yuji Matsumoto: 「Use of XML and Relational Databases for Consistent Development and Maintenance of Lexicons and Annotated Corpora」, LREC 2002, May. 2002.
- [22] 浅原 正幸, 米田 隆一, 山下 亜希子, 伝 康晴, 松本 裕治: 「語長変換を考慮したコーパス管理システム」, 情報処理学会論文誌, Vol.43 No.07, Jul. 2002.
- [23] 浅原 正幸, 松本 裕治: 「形態素解析とチャンキングの組み合わせによる日本語テキスト中の未知語出現箇所同定」, 情報処理学会研究報告 2003-NL-154, Mar. 2003.

付録

A 変更点

A.1 ipadic-2.6.3 から ipadic-2.7.0 への変更点

変更点は<http://chasen.aist-nara.ac.jp/~masayu-a/ipadic/arch/ipadic-2.7.0-diff.txt>にあります。

- 旭化成の水嶋さんから多くの未登録語および誤り事例を報告していただきました。
- 1 b y t e 文字を Onebyte.dic へと別ファイルにした。
- 活用形・活用型定義変更「形容詞・イイ」「特殊・ドス」の導入

A.2 ipadic-2.6.2 から ipadic-2.6.3 への変更点

変更点は<http://chasen.aist-nara.ac.jp/~masayu-a/ipadic/arch/ipadic-2.6.3-diff.txt>にあります。

- 国語研の峯さんから誤り事例を報告していただきました。
- 順天大学の宋さんから誤り事例を報告していただきました。
- 旭化成の水嶋さんから多くの未登録語および誤り事例を報告していただきました。

A.3 ipadic-2.6.1 から ipadic-2.6.2 への変更点

接続表のパラメータを改善致しました。辞書エントリに変更はありません。

A.4 ipadic-2.6.0 から ipadic-2.6.1 への変更点

辞書エントリに変更はありません。

- chasenrc の修正。
- 茶釜 2.3.2 への対応。

A.5 ipadic-2.5.1 から ipadic-2.6.0 への変更点

ipadic-2.5.1 と ipadic-2.6.0 は全く同じ品詞体系を採用しています。

- 長崎大学の鶴丸先生から頂いた未登録語(約 6000 語)を登録した。ここに記して感謝致します。
変更点は<http://chasen.aist-nara.ac.jp/~masayu-a/ipadic/arch/ipadic-2.6.0-diff.txt>にあります。
- その他追加した単語は<http://chasen.aist-nara.ac.jp/~masayu-a/ipadic/arch/ipadic-2.6.0-add.txt>にあります。

- 茶筌 2.3.1 への対応 .
 - .pat, .ary 生成部分の削除
 - ダブル配列辞書 dadic への対応
 - 文字コード変更オプション -i への対応

A.6 ipadic-2.5.0 から ipadic-2.5.1 への変更点

ipadic-2.5.0 と ipadic-2.5.1 は全く同じ品詞体系を採用しています .

- 「助詞-格助詞-連語」相当語句を別ファイル Postp-col.dic にした .
- 一段動詞や助動詞周りのパラメータの変更を行なった .
- 新規登録語のリストは <http://chasen.aist-nara.ac.jp/~masayu-a/ipadic/arch/ipadic-2.5.1-newword.txt> にあります .

A.7 ipadic-2.4.X から ipadic-2.5.0 への変更点

ipadic-2.4.X と ipadic-2.5.0 は全く同じ品詞体系を採用しています .

B 辞書の著作権

本辞書は ICOT の研究成果物を含んでいます .

二次配布の際には , 以下の文面を添付するようにしてください .

Copyright ©copyright 2000, 2001, 2002, 2003 Nara Institute of Science and Technology. All Rights Reserved.

Use, reproduction, and distribution of this software is permitted. Any copy of this software, whether in its original form or modified, must include both the above copyright notice and the following paragraphs.

Nara Institute of Science and Technology (NAIST), the copyright holders, disclaims all warranties with regard to this software, including all implied warranties of merchantability and fitness, in no event shall NAIST be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortuous action, arising out of or in connection with the use or performance of this software.

A large portion of the dictionary entries originate from ICOT Free Software. The following conditions for ICOT Free Software applies to the current dictionary as well.

Each User may also freely distribute the Program, whether in its original form or modified, to any third party or parties, PROVIDED that the provisions of Section 3 ("NO WARRANTY") will ALWAYS appear on, or be attached to, the Program, which is distributed substantially in the same form as set out herein and that such intended distribution, if actually made, will neither violate or otherwise contravene any of the laws and regulations of the countries having jurisdiction over the User or the intended distribution itself.

NO WARRANTY

The program was produced on an experimental basis in the course of the research and development conducted during the project and is provided to users as so produced on an experimental basis. Accordingly, the program is provided without any warranty whatsoever, whether express, implied, statutory or otherwise. The term "warranty" used herein includes, but is not limited to, any warranty of the quality, performance,

merchantability and fitness for a particular purpose of the program and the nonexistence of any infringement or violation of any right of any third party.

Each user of the program will agree and understand, and be deemed to have agreed and understood, that there is no warranty whatsoever for the program and, accordingly, the entire risk arising from or otherwise connected with the program is assumed by the user.

Therefore, neither ICOT, the copyright holder, or any other organization that participated in or was otherwise related to the development of the program and their respective officials, directors, officers and other employees shall be held liable for any and all damages, including, without limitation, general, special, incidental and consequential damages, arising out of or otherwise in connection with the use or inability to use the program or any product, material or result produced or otherwise obtained by using the program, regardless of whether they have been advised of, or otherwise had knowledge of, the possibility of such damages at any time during the project or thereafter. Each user will be deemed to have agreed to the foregoing by his or her commencement of use of the program. The term "use" as used herein includes, but is not limited to, the use, modification, copying and distribution of the program and the production of secondary products from the program.

In the case where the program, whether in its original form or modified, was distributed or delivered to or received by a user from any person, organization or entity other than ICOT, unless it makes or grants independently of ICOT any specific warranty to the user in writing, such person, organization or entity, will also be exempted from and not be held liable to the user for any such damages as noted above as far as the program is concerned.