

Ipadic version 2.7.0 用户手册

浅原正幸 松本裕治

2003 年 11 月

Copyright © 2003 奈良先端科学技术大学院大学情报科学研究科自然言语处理学讲座

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.

该说明书对应 IPADIC 2.7.0 版本。

在标明版权和标注许可文字的条件下，允许全部翻印并发布此说明书。

在全部翻印这一前提下，只要将翻印结果与以上许可文字一同发布，即允许翻印并发布此说明书的修订版。

在上述关于修订版的条件下，允许翻印并发布此说明书的翻译版本。

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

目录

1	安装	2
1.1	UNIX 下辞典的安装	2
1.2	Windows 下辞典的安装	2
2	文件的格式	3
2.1	词性定义文件的词性定义	3
2.2	活用类型定义文件的格式	3
2.3	活用形定义文件的格式	3
2.4	辞典文件格式	4
2.5	连接表文件的格式	6
3	chasenrc 文件	7
4	语素条目的追加	12
4.1	文件的编辑	12
4.2	系统辞典再编译 (UNIX 版)	12
4.3	用户辞典的再编译 (UNIX 版)	13
4.4	辞典的再编译 (Windows 版)	13
5	IPA 词性体系	13
5.1	名词	14
5.1.1	名词-一般	14
5.1.2	名词-固有名词-一般	14
5.1.3	名词-固有名词-人名-一般	14
5.1.4	名词-固有名词-人名-姓	14
5.1.5	名词-固有名词-人名-名	15
5.1.6	名词-固有名词-组织	15
5.1.7	名词-固有名词-地域-一般	15
5.1.8	名词-固有名词-地域-国家	15
5.1.9	名词-代名词-一般	15
5.1.10	名词-代名词-缩略	15
5.1.11	名词-可能副词	15
5.1.12	名词-サ变接续	15
5.1.13	名词-形容动词词干	16
5.1.14	名词-ナイ形容词词干	16
5.1.15	名词-数词	16
5.1.16	名词-附属-一般	16
5.1.17	名词-附属-可能副词	16
5.1.18	名词-附属-助动词词干	16
5.1.19	名词-附属-形容动词词干	17
5.1.20	名词-特殊-助动词词干	17
5.1.21	名词-接尾-一般	17
5.1.22	名词-接尾-人名	17
5.1.23	名词-接尾-地域	17
5.1.24	名词-接尾-サ变接续	17
5.1.25	名词-接尾-助动词词干	17
5.1.26	名词-接尾-形容动词词干	17
5.1.27	名词-接尾-副词可能	18
5.1.28	名词-接尾-助数词	18
5.1.29	名词-接尾-特殊	18
5.1.30	名词-接续词性	18
5.1.31	名词-动词附属性	18
5.2	接头词	18
5.2.1	接头词-名词接续	18
5.2.2	接头词-数词接续	18

5.2.3	接头词-动词接续	18
5.2.4	接头词-形容词接续	19
5.3	动词	19
5.3.1	动词-独立 力变 【活用形】	20
5.3.2	动词-附属 力变 【活用形】	20
5.3.3	动词-独立 サ变・スル 【活用形】	20
5.3.4	动词-独立 サ变・ースル 【活用形】	21
5.3.5	动词-独立 サ变・ーズル 【活用形】	21
5.3.6	动词-独立 一段 【活用形】	21
5.3.7	动词-附属 一段 【活用形】	21
5.3.8	动词-接尾 一段 【活用形】	21
5.3.9	动词-独立 五段・力行イ音变 【活用形】	21
5.3.10	动词-附属 五段・力行イ音变 【活用形】	21
5.3.11	动词-附属 五段・力行促音变 【活用形】	22
5.3.12	动词-附属 五段・力行促音变 【活用形】	22
5.3.13	动词-独立 五段・ガ行 【活用形】	22
5.3.14	动词-独立 五段・サ行 【活用形】	22
5.3.15	动词-附属 五段・サ行 【活用形】	22
5.3.16	动词-独立 五段・タ行 【活用形】	22
5.3.17	动词-独立 五段・ナ行 【活用形】	22
5.3.18	动词-独立 五段・バ行 【活用形】	22
5.3.19	动词-独立 五段・マ行 【活用形】	22
5.3.20	动词-附属 五段・マ行 【活用形】	22
5.3.21	动词-独立 五段・ラ行 【活用形】	22
5.3.22	动词-附属 五段・ラ行 【活用形】	23
5.3.23	动词-接尾 五段・ラ行 【活用形】	23
5.3.24	动词-独立 五段・ラ行特殊 【活用形】	23
5.3.25	动词-附属 五段・ラ行特殊 【活用形】	23
5.3.26	动词-独立 五段・ワ行ウ音变 【活用形】	23
5.3.27	动词-附属 五段・ワ行ウ音变 【活用形】	23
5.3.28	动词-独立 五段・ワ行促音变 【活用形】	23
5.3.29	动词-附属 五段・ワ行促音变 【活用形】	23
5.3.30	动词-独立 四段・ハ行 【活用形】	24
5.3.31	动词-独立 ラ变 【活用形】	24
5.3.32	动词-独立 上二段・ハ行 【活用形】	24
5.3.33	动词-独立 下二段・ア行 【活用形】	24
5.3.34	动词-独立 一段・得ル 【活用形】	24
5.4	形容词	24
5.4.1	形容词-独立 形容词・アウオ段 【活用形】	25
5.4.2	形容词-独立 形容词・イ段 【活用形】	26
5.4.3	形容词-独立 形容词・イイ段 【活用形】	26
5.4.4	形容词-独立 形容词・不变化型 【活用形】	26
5.4.5	形容词-附属 形容词・アウオ段 【活用形】	26
5.4.6	形容词-附属 形容词・イ段 【活用形】	26
5.4.7	形容词-附属 形容词・イイ段 【活用形】	26
5.4.8	形容词-附属 形容词・不变化型 【活用形】	26
5.4.9	形容词-接尾 形容词・アウオ段 【活用形】	26
5.4.10	形容词-接尾 形容词・イ段 【活用形】	26
5.5	副词	26
5.5.1	副词-一般	26
5.5.2	副词-助词类接续	26
5.6	连体词	27
5.6.1	连体词	27
5.7	接续词	27
5.7.1	接续词	27
5.8	助词	27
5.8.1	助词-格助词-一般	27

5.8.2	助词-格助词-引用	27
5.8.3	助词-格助词-连词	27
5.8.4	助词-接续助词	27
5.8.5	助词-系助词	27
5.8.6	助词-副助词	28
5.8.7	助词-并列助词	28
5.8.8	助词-终助词	28
5.8.9	助词-副助词／并列助词／终助词	28
5.8.10	助词-连体化	28
5.8.11	助词-副词化	28
5.8.12	助词-特殊	28
5.8.13	助词-感动助词	28
5.9	助动词	29
5.9.1	助动词 五段・ラ行アル【活用形】	29
5.9.2	助动词 五段・ラ行ゴザル【活用形】	29
5.9.3	助动词 形容词・イ段【活用形】	29
5.9.4	助动词 特殊・ナイ【活用形】	29
5.9.5	助动词 特殊・タ【活用形】	29
5.9.6	助动词 特殊・ダ【活用形】	29
5.9.7	助动词 特殊・デス【活用形】	29
5.9.8	助动词 特殊・ドス【活用形】	29
5.9.9	助动词 特殊・ジヤ【活用形】	29
5.9.10	助动词 特殊・マス【活用形】	30
5.9.11	助动词 特殊・ヌ【活用形】	30
5.9.12	助动词 特殊・ヤ【活用形】	30
5.9.13	助动词 不变化型【活用形】	30
5.9.14	助动词 文言・??【活用形】	30
5.10	感动词	30
5.10.1	感动词	30
5.11	标点	30
5.11.1	标点-普通	30
5.11.2	标点-字母	30
5.11.3	标点-句号	30
5.11.4	标点-逗号	31
5.11.5	标点-空格	31
5.11.6	标点-左括号	31
5.11.7	标点-右括号	31
5.12	补白	31
5.12.1	补白	31
5.13	其他	31
5.13.1	其他-感动	31

参考文献	31
------	----

附录	33
----	----

A 变更点	33
-------	----

A.1	从 ipadic-2.6.3 到 ipadic-2.7.0 的变更点	33
A.2	从 ipadic-2.6.2 到 ipadic-2.6.3 的变更点	33
A.3	从 ipadic-2.6.1 到 ipadic-2.6.2 的变更点	33
A.4	从 ipadic-2.6.0 到 ipadic-2.6.1 的变更点	33
A.5	从 ipadic-2.5.1 到 ipadic-2.6.0 的变更点	33
A.6	从 ipadic-2.5.0 到 ipadic-2.5.1 的变更点	33
A.7	从 ipadic-2.4.X 到 ipadic-2.5.0 的变更点	34

B 辞典的版权	34
---------	----

卷首语

语素分析系统 ChaSen 是奈良先端科技大学院大学作为自由软件发布的语言处理系统。本辞典是对 chasen (2.3.2 版本以后) 所使用的日语词典 (ipadic2.7.0) 的说明。本辞典在信息处理振兴事业协会 (IPA) 设定的 IPA 词性体系 (THiMCO97) 基础上作了部分修改。在取得使用许可后, 本辞典对刊登在信息处理开发机构 (RWCP) 发布的“文本数据库报告 (1998 年)”上的 IPA 词性体系 (THiMCO97) 的说明进行了摘录和部分修改。

目前的 IPA 词性体系日语辞典是在以奈良先端科技大学院大学情报科学研究科的鹿野清宏教授为代表的“日语听写基本软件的开发”(关于 IPA 独创性先进信息技术的研究开发) 小组对 1998 年 5 月发布的 IPA 词性体系日语辞典 (ipadic1.0b2) 进行了大幅修订和改进之后, 经过各种各样的修订而成。

在此, 对参与建立本辞典系统的各位表示衷心感谢。

关于本说明书有任何疑问, 请与以下地址联系:

〒 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.naist.jp

我们通过以下地址提供最新信息:

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

1 安装

1.1 UNIX 下辞典的安装

本辞典要求 ChaSen-2.3.2 以上环境。请先下载并安装 ChaSen 系统。
一般的安装方法：

1. 执行./configure 命令

```
%./configure
```

安装目录是 ChaSen 自动设定的，若要手动改变安装目录，请执行以下命令：

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

辞典就会生成在/home/masayu-a/ipadic 下。

2. 执行 make 命令

```
% make
```

使用操作系统的标准的 make，有时会出现编译失败。此时可使用 GNU make。

3. 在 root 权限下执行 make install

```
# make install
```

默认安装在/usr/local/share/chasen/dic/ipadic 下(环境不同时多少会有差别)。在用户目录下安装时，不需要 root 权限。

4. 编辑/usr/local/etc/chasenc

第一次安装 ChaSen 且第一次安装辞典时，安装文件会自动生成在/usr/local/etc/chasenc 下。如果未自动生成，请一边查看程序包中生成的 chasenc，同时编辑/usr/local/etc/chasenc。

1.2 Windows 下辞典的安装

以下说明假设 wincha 是以下列形式安装的：

```
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
```

安装 WinCha 时 IPADIC 会自动安装。但若用户要对辞典进行手动加工，则需要 SJIS 的辞典程序包。SJIS 辞典程序包可以从此下载：

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

将展开的辞典文件(Noun.dic 等，以 extension.dic 结尾的文件)，连接表文件(connect.cha)，词性定义文件(grammar.cha)，活用类型定义文件(ctypes.cha)，活用形定义文件(cforms.cha)等拷贝到安装了 Windows 版 ChaSen 的目录中的c:\Program Files\chasen21\dic目录下。

把辞典程序包中的 Makefile.bat 拷贝到c:\Program Files\chasen21并在命令提示符中执行它。

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

在 Windows XP/2000/NT 环境下，必须在 Administrator 权限下安装。

2 文件的格式

2.1 词性定义文件的词性定义

格式文件 grammar.cha 描述的是词性列表。活用词性<最高级词性定义>标记 %。对于活用词性，必须在 ctypes.cha 列举可能的活用类型，在 cforms.cha 中描述可能的活用形。

(接头词
(名词接续)
(动词接续)
(形容词接续)
(数词接续))

(动词 %
(独立)
(附属)
(接尾))

- <词性定义> ::= “(<最高级词性信息>(<低级词性信息>)*)”
- <最高级词性定义> ::= <最高级词性名称> | “<最高级词性名称>%”
- <低级词性定义> ::= <词性分类名称> | “<词性分类名称>(<低级词性信息>)*”

2.2 活用类型定义文件的格式

活用类型定义文件 ctypes.cha 描述了各个词性可以使用哪些活用类型。

((形容词独立)
(形容词・アウオ段
形容词・イ段
不变化型)
)

- <活用类型定义> ::= “((<词性名>)(<活用类型>*))”

2.3 活用形定义文件的格式

活用形定义文件 cforms.cha 描述了各活用类型采用哪种活用形，以及各活用形可以采用哪种活用词尾。活用词尾可描述汉字形、假名形和发音形。

(形容詞・イ段		
(; (詞干	*)
(基本形	い	イ)
(文言基本形	*	*)
(未然又接続	から	カラ)
(未然ウ接続	かる	カロ)
(连用タ接続	かつ	カツ)
(连用テ接続	く	ク)
(连用テ接続	くっ	クッ)
(连用ゴザイ接続	ゆう	ユウ ユー)
(连用ゴザイ接続	ゆう	ユウ ユー)
(体言接続	き	キ)
(假定形	けれ	ケレ)
(命令 e	かれ	カレ)
(假定縮略 1	けりゃ	ケリ ャ)
(假定縮略 2	きゃ	キ ャ)
(ガル接続	*))
)		

- <活用形定義> ::= “(<活用类型名称>(<活用形信息>*))”
- <活用形信息> ::= “(<活用形><汉字形活用词尾>< 假名形活用词尾>< 发音活用词尾>)” | “(<活用形名称><汉字形活用词尾><假名形活用词尾>)” | “(<活用形名称><汉字形活用词尾>)”

2.4 辞典文件格式

以下是辞典文件的示例。辞典文件按各词性分类。

```
(词性 (名词 普通)) ((词条 (お正月 3641)) (训读 オシヨウガツ) (发音 オシ ヨーガツ))

(词性 (动词 独立)) ((词条 (あきらめる 2377)) (训读 アキラメル) (活用类型一段))

(词性 (名词 普通)) ((词条 (天文学 3556)) (训读 テンモンガク)
(复合词
((词性 (名词 普通)) (词条 天文) (训读 テンモン))
((词性 (名词 接尾 普通)) (词条 学) (训读 ガク)) ))
```

辞典文件中的语素条目定义如下：

- <语素词目> ::= “(<词性信息>(<词条信息><语素信息>*))”
- <词性信息> ::= “(词性(<词性名>))”
- <词条信息> ::= “(词条(<词条><语素生成成本>))” | “(词条<词条>)”
- <语素信息> ::= <训读信息> | <发音信息> | <活用类型信息> | <附加信息> | <词义信息> | <复合词信息>
- <训读信息> ::= “(训读<训读>)”
- <发音信息> ::= “(发音<发音>)”
- <活用类型信息> ::= “(活用类型<活用类型>)”
- <复合词信息> ::= “(复合词<构成词词目>*)”
- <构成词词目> ::= “(<词性信息><词条信息><构成词语素信息>*)”
- <构成词语素信息> ::= <训读信息> | <发音信息> | <活用类型信息> | <附加信息> | <词义信息> | <活用形信息>

- <活用形信息>::=“(活用形<活用形>)”

需要说明的是，<语素信息>及<构成词语素信息>中出现的项目不允许重复。

- <词性名>

词性名。阶梯结构。各层之间以空格隔开。

例:

(词性 (名词 普通))
(词性 (动词 独立))
(词性 (名词 接尾 普通))

- <词条>

它是文章中出现的形式。活用词只收录基本形。

例:

(词条 (お正月 3641))
(词条 (あきらめる 2377))
(词条 (天文学 3556))
(词条 天文)
(词条 学)

- <语素生成成本>

词条旁边的数字称为语素生成成本。数值越小说明该词出现几率越大。IPADIC 的语素生成成本是以从分析完毕的数据中得到的单词出现概率为依据，计算出来的数值。

用户增加语素时，如果预设为某个与其使用频率相近的语素的数值，则多数情况下都能客观地反映在分析中。未能反映时，可调低语素成本的数值。

例:

(词条 (お正月 3641))
(词条 (あきらめる 2377))
(词条 (天文学 3556))

- <读音>

它是假名标记。用片假名描述。

例:

(训读 オシヨウガツ)
(训读 アキラメル)
(训读 テンモンガク)
(训读 テンモン)
(训读 ガク)

- <发音>

它是发音标记。用片假名描述。

例:

(发音 オシ ヨーガツ)

- <活用类型>

活用词都有活用类型。只认可 ctypes.cha 给出了定义的活用类型。

例:

(活用类型 五段・サ行)

- <活用形>

构成复合词的语素是活用词且非基本形的情况下, 用来描述构成词词目。

例:

(活用形 未然ウ接续)

- <附加信息>

单词词目的附加信息。用户可自由使用。描述语调信息以及其他词性体系下的词性名等。

例:

(附加信息 语调型 =4)

- <词义信息>

单词词目的词义信息。用户可自由使用。可描述同义词信息及国語辞典的条目等。

例:

(词义信息 "思い切る。仕方がないと断念する。")

2.5 连接表文件的格式

以下是连接表文件 connect.cha 中连接规则的示例。* 表示适用的所有可能性。适用于文件前方的规则的连接规则, 是根据文件后方的规则书写的。因此, 必须先写一般性规则, 后写详细规则。

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

(( (((动词 独立) 五段・ラ行アル 连用形))
  (((助动词) 特殊・マス )) ) 604)

(( (((助词 接续助词) * * て))
  (((助词 系助词) * * も))
  (((形容词 附属) 形容词・アウオ段 * よい))) 35)
```

- <连接规则词目> ::= “(<连接信息><连接成本>)”
- <连接信息> ::= <词性定义><词性定义>+
- <词性定义>+ ::= <词性定义> | <词性定义>+
- <词性定义> ::= “(<词性信息><活用类型信息><活用形信息><词汇化词性規則>)” | “(<词性信息><活用类型信息><活用形信息>)” | “(<词性信息><活用类型信息>)” | “(<词性信息>)”
- <词性信息> ::= “(<词性名>)”
- <活用类型信息> ::= “<活用类型>” | “*”
- <活用形信息> ::= “<活用形>” | “*”
- <词汇化词性規則> ::= “<词汇化词性定义>” | “*”

3 chasenrc 文件

chasenrc 文件是用来给语素分析程序所必须的各种备选项下定义的。这些定义通常被写在PREFIX/etc/chasenrc, 也可以写在使用者的主目录下的‘.chasenrc’这一文件中。也可以根据启动选项等来指定chasenrc文件。具体来说, 按照以下优先级顺序读取chasenrc文件。

1. 启动时用-r选项指定的文件。
2. 环境变量CHASENRC指定的文件。
3. 用户主目录下的.chasen2rc。
4. 用户主目录下的.chasenrc。
5. PREFIX/etc/chasenrc。

下面将给出设定项目一览。其中, 必须指定“DADIC”, “未知词性”, “词性成本”。

1. 语法文件的目录

指定语法文件(grammar.cha,ctypes.cha,cforms.cha,connect.cha)存在的目录。

(语法文件/usr/local/lib/chasen/ipadic/dic)

“语法文件”可被省略, 此时认为 chasenrc 文件指定了与某目录相同的目录。Chasen 附属词典 ipadic1.01 以后的 chasenrc 文件中, “语法文件”被省略。

2. 系统词典

通过描述除末尾的扩展名之外的文件名来指定双排序词典(chadic.fda,lex,datg)。也可指定多个词典组合而成的词典。此外, 如果描述相对路径, 即不以“/”开头的路径, 则视为指定了与语法文件相同的目录。例如, 如下指定:

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

通过这一叙述, 读取了下列两组词典文件:

- (a) 与语法文件在同一目录下的 chadic.{da,lex,dat}
- (b) 在/home/rikyu/mydic/下的 chadic.{da,lex,dat}

查字典时, 这两组词典都被调用了¹。

为使用 Darts²双排序词典, 指定“DADIC”。

(DADIC chadic)

在以上叙述中, 与语法文件在同一目录下的 chadic.da,chadic.lex,chadic.dat 被读取。

可使用词典的最大数量设定为 32 个。

3. 未知词词性

发现未知词时, 指出将它作为哪类词性来适用连接规则。指定多个词性时, 每个词性分别适用连接规则。

(未知词词性 (名词 ㄅ变接续)) ; 指定一个词性
(未知词词性 (名词 ㄅ变接续) (名词 普通)) ; 指定多个词性

¹虽然在一组词典中不会多次登录相同的语素, 但是在多个词典的检索过程中, 一个语素在多个词典中同是出现的可能性存在。在这种情况下, 系统会得到多次对于同一语素的信息。

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

4. 词性的成本

在语素分析程序中，将分析结果的优先信息作为成本来计算。当分析存在模糊性时，成本总计较低者优先。“词性成本”定义了各词性成本的倍率和“未知词”的成本。成本取正的整数值。

```
(词性成本
  ((*)          1)
  ((未知词)     500)
  ((名词)       2)
  ((名词 固有名词) 3)
)
```

当同一词性的成本定义被多次指定时，最后一次优先。在以上描述中，“名词”的语素成本基本为 2 倍，只有“名词—固有名词”之下再细分的名词语素成本才为 3 倍。词前有 ‘(*)’ 标记，在此没有给出明确定义的语素的成本均为 1 倍（本身的成本值）。未知词的语素成本值均为 500。

5. 连接成本和语素成本的相对权值定义

语素分析中的成本计算是计算语素成本与连接成本的总和。要给这两类成本加上不同的权值，可以指定相对权值。成本的分析结果是把两者各自的成本乘以此处指定的权值再相加。省略时，权值为 1。

```
(连接成本权值 1) ; 默认值
(语素成本权值 1) ; 默认值
```

6. 成本范围

在语素分析过程中，并不总得出成本最低的结果。有时成本在一定范围内变化。可指定这一允许范围。要输出成本范围中所有的解，可使用 -m 选项和 -p 选项。

```
(成本范围 0) ; 默认值
```

成本范围也可用 -w 选项指定。在此情况下，用 -w 选项指定者优先。

7. 未定义连接成本的定义

指定的是连接规则文件中未对其连接规则进行定义的语素之间的连接成本。若不设定未定义连接成本，或设定为 0，则表示连接规则未定义的语素间决不连接。默认值为 0。

```
(未定义连接成本 500)
```

8. 输出格式

通过指定输出格式，可改变分析结果的输出形式。

```
(输出格式 "%m\t%y\t%P-\n")
```

输出格式也可用 -F 选项指定。在此情况下，用 -F 选项指定者优先。详情参照第 3 节。

9. BOS 字符串

指定显示在分析结果文章开头的字符串。使用 “%S” 则显示分析文章全文。默认为空字符串(不显示任何东西)。

```
(BOS 字符串 "分析文章: [%S]\n")
```

10. EOS 字符串

指定显示在分析结果文章末尾的字符串。使用 “%S” 则显示分析文章全文。默认为 “EOS\n”。

```
(EOS 字符串 "文末\n")
```

11. 空白词性

chasen 将半角空格文字(ASCII 32)和 Tab 键(ASCII 9)视为空格, 在分析时忽略。一般情况下, 分析结果中不输出空格信息, 但通过设定“空白词性”, 就能输出有关空格的信息。例如, 进行如下设定, 则空格输出为“标点-空格”。

```
(空格词性 (标点 空格))
```

此外, 如将输出格式设为“%m”, 且指定空格词性(任何词性均可), 则可输出与分析文章完全相同的内容。

12. 注释

可将以某字符串开始并以某字符串结束的字符串作为注释处理, 忽略该字符串的部分内容进行分析。在分析结果中, 该字符串输出为一个语素。

chasenrc 文件描述由开始字符串和结束字符串组成的列表以及输出时的词性名称或格式字符串。结束字符串可省略。在此情况下, 与开始字符串相一致的字符串本身将作为注释处理。此外, 如果省略输出时的词性名称或格式字符串, 那么有关语素的所有信息都不会被输出。

```
(注释 (( "<" ">" ) "%m\n")
      (( "[" "]" ) (标点 普通))
      (( "(" ")" ) (标点 普通))
      (( "\"" "\"" ) (名词 引用字符串))
      (( "[" "]" ) ))
)
```

例如, 作上述描述, 则分析、输出如下:

- `` 这样的以“<”开始以“>”结束的字符串输出为它本身。
- “[”或“]”输出为“标点-普通”。
- “`hello(again)`”这样用双重引用符号括起来的字符串输出为“名词-引用字符串”。
- 像“[ちゃん]”这样以“[”开始以“]”结束的字符串在分析会忽略, 分析结果中不输出该字符串的信息。

13. 连接词性

用于某一词性的语素连着出现时, 将它连接为一个语素来输出。

```
(连接词性 ((复合名词) (名词) (接头词 名词接续) (接头词 数词接续))
           ((标点)))
```

例如, 上述描述中, 词性将像下面这样连接:

- 把连着出现的“名词”“接头词-名词接续”“接头词-数词接续”连接起来, 显示为“复合名词”。必须先把“复合名词”写入词性定义文件 grammar.cha 中。
- 把接着出现的“标点”连起来, 显示为“标点”。

14. 复合词输出

对于语素词典文件(.dic)中给出定义的复合词, 可以选择输出整个复合词的语素信息(“复合词”), 或输出构成复合词的单个单词的语素信息(“构成词”)。默认为“复合词”。

```
(复合词输出 "复合词")
```

另外, 输出复合词时, 可用-Oc, -Os 选项控制。

15. 分割符号

把运行-j 选项时的文章的句读排列成一个字符串(参照??节)。全角字符和半角字符两种句读均可使用。例如：

(区切り文字".、,!? .,!? ")

进行这一定义,则全角字符的「.、,!?」中的任意一个,或半角字符“.,!? ”(注意有空格文字)中的任意一个符号成为文章的句读。

输出的格式

通过使用-F选项,或者在 chasenrc 文件的[输出格式]中指定输出的格式,可以改变解析结果的输出形式。

输出格式的文字串末尾如果有'\n',则输出有改行显示,并且在文末会自动添加'EOS'。如果输出格式的末尾没有'\n',则所有的输出都会在一行显示,在文末才会有改行。

此外,在输出格式中指定'-f'、'-e'、'-c'时,则分别按照-f、-e、-c的格式显示。

这里列出几个输出格式的例子:

- 和默认格式(-f)一样的输出
"%m\t%y\t%M\t%U(%P-)\t%T\t%F\t\n"或者"-f"
- 单词、读音、词性用 tab 分割表示
"%m\t%y\t%P-\n"
- 只显示单词
"%m\n"
- 显示单词间的分割(用空白区分单词)
"%m_"
- 汉字假名变换
"%y"
- 注音表示。“漢字(かな)”的表示形式
"%x_()"

输出格式的变换文字如下表所示:

变换文字	功能
%m	单词(出现形)
%M	单词(基本形)
%y, %y1	读音的第一候补(出现形)
%Y, %Y1	读音的第一候补(基本形)
%y0	全部读音(出现形)
%Y0	全部读音(基本形)
%a	发音的第一候补(出现形)
%A	发音的第一候补(基本形)
%a0	全部发音(出现形)
%A0	全部发音(基本形)
%rABC	附带注音的表示(如: “A漢字BかなC”)(注 1)
%i, %i1	附加信息的第一候补
%i0	附加信息的全体
%Ic	附加信息(空文字列表示, 或者如果是“NIL”的话, 用文字c表示)(注 1)
%Pc	用文字c来区分各级词性的文字列
%Pnc	用文字c来区分 1 n(n:1 9)级词性的文字列
%h	词性的编号
%H	词性文字列
%Hn	第n(n:1 9)级的词性(如果没有, 则是最深一级的词性)
%b	0(仅是与旧版本的互换性)
%BB	词性的细分类(如果没有, 则为该词性)
%Bc	词性的细分类(如果没有, 则为文字c)(注 1)
%t	活用型的编号
%Tc	活用型(如果没有, 则为文字c)(注 1)
%f	活用形的编号
%Fc	活用形(如果没有, 则为文字c)(注 1)
%c	语素的计算成本
%S	解析全句
%pb	用“*”表示最优路径, 用“_”表示其他路径
%pi	路径的编号
%ps	路径的语素的开始位置
%pe	路径的语素的终止位置 +1
%pc	路径的计算成本
%ppiC	列举出用文字C区分的前方连接路径的编号
%ppcC	列举出用文字C区分的前方连接路径的计算成本
%%B/STR1/STR2/	用STR1表示词性的细分类, 用STR2表示其他(注 2)
%%I/STR1/STR2/	附加信息不为“NIL”或者“”(空文字列)时, 用STR1表示; 否则用STR2表示(注 2)
%%T/STR1/STR2/	是活用时, 用STR1表示; 否则用STR2表示(注 2)
%%F/STR1/STR2/	与%%T/STR1/STR2/相同
%%U/STR1/STR2/	是未知词的话, 用STR1\表示; 否则用STR2表示(注 2)
%%U/STR/	是未知词的话, 用“未知词”表示; 否则用STR2表示(与%%U/未知词/STR/相同)(注 2)
%%	符号%
.	指定字段的长度
-	指定字段的长度
1-9	指定字段的长度
\n	换行符
\t	制表符
\\	符号\
\'	符号'
\"	符号"

注 1 ipadic 中, 当一个语素有多个读音的情况时(如[行く(いく/ゆく)]), 我们用半角的括号和斜线来表示它的读音, 如[イ/ユク]。一般的情况下, 在读音的输出(输出格式%y中, 只显示读音的第一候补[イク], 使用%y0的输出格式时, 则可显示全部的读音[イ/ユク])。

注 1 当A、B、C、c为空白文字时, 不显示任何内容。

注 2 在‘/’处，可以使用任意的文字。此外，也可以使用括号“(){}[]<>”。示例如下：

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

4 语素条目的追加

4.1 文件的编辑

下载 ipadic-X.X.X.tar.gz 或 ipadic-sjis-X.X.X.zip 并解压缩。下载文件可在此找到：

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

按照上述格式编辑。

- *.dic
语素辞典
- connect.cha
连接表
- grammar.cha
词性定义文件
- ctypes.cha
活用类型定义文件
- cforms.cha
活用形定义文件

4.2 系统辞典再编译（UNIX 版）

要编辑系统的语素辞典比如更改词性体系等，必须对辞典进行再编译。

1. 执行./configure

要手动改变安装目录，请执行以下操作：

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

2. 执行 make

```
% make
```

使用操作系统的标准的 make，有时会出现编译失败。此时可使用 GNU make。

3. 在 root 权限执行 make install

```
# make install
```

默认安装在/usr/local/share/chasen/dic/ipadic 下（环境不同时多少会有差别）。

在 user 目录下安装时，不需要 root 权限。

4.3 用户辞典的再编译（UNIX 版）

词性体系不发生改变的情况下增加简单词汇时，可使用用户辞典。首先建立用户辞典专用的目录。增加扩展名为.dic 的语素辞典文件后，输入以下命令：

```
% mkdir ~/mydic
% cd ~/mydic
% emacs Noun2.dic
(描述语素信息)
$ 'chasen-config --mkchadic'/makeda -i e chadic *.dic
```

在 makeda 中用 i 选项指定文字编码格式。可指定以下 4 种编码格式。（e:EUC-JP, s:Shift JIS, w:UTF-8, a:ISO-8859-1）。

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

然后把 chasenrc 文件以.chasenrc 文件名拷贝到 home 目录下。

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

编辑.chasenrc，如下图所示指定“语法文件”，把用户辞典增加到“DADIC”中。

```
(语法文件/usr/local/share/chasen/dic/ipadic)
(DADIC chadic
    /home/masayu-a/mydic/chadic)
```

4.4 辞典的再编译（Windows 版）

将展开的文件拷贝到安装了 Windows 版 chasen 的目录中的 dic 这个目录下。描述单词完毕后，在命令提示符中执行 Makefile.bat。

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

5 IPA 词性体系

说明格式

词性名

以下，有时把词性名称为“标记”。对各个词性进行说明时，用以下符号加以注释。

词性的解说

例：单词词例

* 关于词性解说的参考事项

& 关于读音、活用形的参考事项

有关词性名的注意事项

本日语辞典建立在 IPA 词性体系（THiMCO97）基础上，编写 ChaSen 辞典时，我们对词性体系进行了几处变更。现将本词性体系的特征和变更点总结如下：

- 词性分成多个层次。例如，“名词固有名词人名姓”是由四个层次组成的词性名。以下用连字符分隔，表示为“名词-固有名词-人名-姓”。在 ChaSen2.0 中，由于能对任意层数的词性层次进行定义，因此可直接将其定义为语法文件（grammar.cha）。
- 在 THiMCO97 中，像“动词一段连用形独立”这样，定义时词性的分类和活用类型、活用形是混合在一起的。而在 ChaSen 中，由于词性分类的定义和有关活用的定义是相互分开的，因此把它分成三个项目（词性名、活用类型、活用形），描述成“动词—独立—一段—连用性”。
- 我们根据以下标准，对用来给词性名下定义的分类名称进行了变更：
 1. 将“（助动词词干）”“（形容动词词干）”这些带圆括号的名称中的圆括号去掉。
 2. 定义为“动词接尾词（助动词）”“形容动词接尾词（助动词）”的，由于“（助动词）”部分属多余，因此将它省略，写成“动词-接尾”“形容动词-接尾”。
 3. 动词分成“动词”“动词附属”“动词接尾”这几个大类。由于在 chasen 的词性层次定义中，“动词”这一描述代表所有动词，因此，为了方便区分，增加“独立”这一小类，变为“动词-独立”，“动词-附属”“动词-接尾”这几类。
同样地，关于除活用词以外的单词的词性名，把“名词”“名词固有名词”“名词固有名词人名”“名词固有名词人名姓”这样的分类增加“普通”这一类别，变为“名词-普通”“名词-固有名词-普通”“名词-固有名词-人名-普通”“名词-固有名词-人名-姓”，进行排他性的词性定义。
 4. 关于用言活用形，THiMCO97 体系根据其后续助动词的种类进行了非常详细的定义，如“未然ナイ接续”“未然レル接续”“未然ウ接续”“连用タ接续”“连用マス接续”“连用タイ接续”等等。但对各种活用类型而言，有“未然”“连用”这样不同的词尾的则很少。因此，我们在定义活用形的名称时把“未然形”“连用形”“基本形”“假定形”“命令形”等作为基本的活用形，只有对特殊的活用形，才使用 THiMCO97 的活用形名称。此外，由于 ChaSen 把辞典中出现的形定为“基本形”，因此，我们把 THiMCO97 中的“词条形”这一活用形名称改为“基本形”这一名称。
 5. 在 THiMCO97 体系中，“未然ウ接续”是五段活用动词后接助动词“う”，其他活用类型的动词后接“よう”这一活用。在此，仅把“う”视为单词（助动词），而“来よ（う）”“食べよ（う）”这些词中的“よう”一律视为活用词尾。
- 在 ipadic2.0 以后的辞典中，单词添加了“发音”这一字段。这是“日语听写基本软件会开发”小组添加的。例如，系助词“は”的读音为“ワ”，“常識”的读音为“ジョーシキ”，长音用“ー”表示。另外，对于拼写和词性都相同，只有读音不同的词，如“私（ワタシ/ワタクシ）”，则标出所有可能的读音：“ワタシ/ワタクシ”，收录为一个词。

5.1 名词

5.1.1 名词-一般

普通名词或未进行下一级分类的名词。

5.1.2 名词-固有名词-一般

普通固有名词或未进行下一级分类的固有名词。

5.1.3 名词-固有名词-人名-一般

姓和名不分的人名，外国人名，或姓、名无法确定的人名。

例：「お市の方」...

5.1.4 名词-固有名词-人名-姓

主要是日本人的姓。

例：「山田」...

5.1.5 名词-固有名词-人名-名

主要是日本人的名字。

例:「太郎」...

5.1.6 名词-固有名词-组织

表示组织名称。

例:「通産省」「NHK」...

5.1.7 名词-固有名词-地域-一般

表示国名以外的地名。

例:「アジア」「バルセロナ」「京都」...

5.1.8 名词-固有名词-地域-国家

国家名称。

例:「日本」「オーストラリア」...

5.1.9 名词-代名词-一般

即代名词，不定代词。

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 名词-代名词-缩略

代名词和系助词「は」相连组成的词的缩略形式<口语>。

例:「ありゃ」「こりゃ」「こりゃあ」「そりゃ」「そりゃあ」...

5.1.11 名词-可能副词

星期、月等表示时间的有副词性用法的名词。表示数量和比例等可作为副词使用的名词。

例:「金曜」「一月」「午後」「少量」...

* 在原来的 IPA 词性体系，“名词—可能副词”中，把实际作为副词使用的名词称作“名词可能副词副词性”，这些词可以作为副词使用；把不作为副词使用的词称作“名词可能副词”。在本体系中，不论文中用法如何，一律把可以作副词使用的词称为“名词—可能副词”。

5.1.12 名词-サ变接续

采用格要素，可以后接「する」「できる」「なさる」「くださる」等的词。

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)」归为“副词-助词类接续”

* 考虑到「名词-普通」「名词-サ变接续」的多重用法，优先后者定义词性名。

5.1.13 名词-形容词词干

形容词词干，即出现在“な”之前的词。

例：「健康」「安易」「駄目」「だめ」...

* 原来的 IPA 词性体系中称为「名词（形容词词干）」，本体系将第 2 层的「（形容词词干）」的括弧去掉。

* 考虑到「名词-普通」「名词-形容词词干」的多重用法，优先后者命名词性名。但是，自然一词在作为 nature 之意使用的「自然」和作为「自然な」之意使用时，意思不同(前者的连体用法为「自然の」)，因此把前者定义为「名词-普通」、后者定义为「名词-形容词词干」，分别收录。

5.1.14 名词-ナイ形容词词干

出现在助动词「ない」前，起形容词作用的词。

例：「申し訳」「仕方」「とんでも」「違い」...

* 在原来的 IPA 词性体系中被视为形容词。由于有「申し訳-ない」「申し訳-ありません」「申し訳-ございせん」等派生，因此统一作为词干看待。但并非所有归类到「ナイ形容词词干」中的词都有这样的用法。

5.1.15 名词-数词

汉数字，阿拉伯数字以及“何（回）”“数（％）”“幾（夜）”。

例：「0」「1」「2」「何」「数」「幾」...

5.1.16 名词-附属-一般

接在连体词，「の（格助词）」，活用词基本形后使用的词中，不适用以下低级分类中的词。包括所谓的形式名词。

* 如果作为普通名词使用，即使有连体修饰，也不是「附属」词。

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 名词-附属-可能副词

接在连体词，「の（格助词）」，活用词基本形后使用的词中，可起副词作用的词。

* 在上下文中实际起副词作用的，IPA 词性体系在其后附加[副词的]这一说明。本体系则省略。

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」「っきり」「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」「俣」「みぎり」「矢先」...

5.1.18 名词-附属-助动词词干

接在连体词，「の（格助词）」，活用词基本形后使用的词中，学校语法视为助动词的「ようだ」的词干部分。

例：「よう」「やう」「様(よう)」

* 在原来的 IPA 体系中，写作「名词-附属-（助动词词干）」。

5.1.19 名词-附属-形容动词词干

接在连体词,「の(格助词)」,活用词基本形后使用的词中,可与“な(助动词「だ」)的体言接续”相接续的词。

例:「みたい」「ふう」...

* 在原来的 IPA 体系中,写作「名词附属(形容动词词干)」。

5.1.20 名词-特殊-助动词词干

接在基本形后使用的,学校语法视为助动词的「そうだ(传闻)」的词干部分。

例:「そう」...

* 在原来的 IPA 体系中,写作「名词特殊(助动词词干)」。

5.1.21 名词-接尾-一般

与名词或者其他词性的词干[ガル接续]和[连用タイ接续]相连组成复合名词的词中,不适用下一级分类的词。一般来说,比“接尾词”范围更广,一般作为复合名词的后半部分使用。

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」「类」「面」「用」...

5.1.22 名词-接尾-人名

名词中,专门与人名相连组成名词的接尾词。

例:「君」「様」「著」等。

5.1.23 名词-接尾-地域

名词中,专门与地名相连组成名词的接尾词。

例:「町」「市」「県」等

5.1.24 名词-接尾-サ变接续

与名词相连组成名词的接尾词中,可后接「スル」的词。

例:「化」「視」「分け」「入り」「落ち」「買い」...

5.1.25 名词-接尾-助动词词干

与其他词性的连用形相连接,学校语法视为助动词词干的「そうだ(様態)」的词干部分。

例:「そう」

* 在原来的 IPA 体系中,写作「名词接尾(助动词词干)」。

5.1.26 名词-接尾-形容动词词干

与其他名词和活用词连用形相连接的接尾词中,后接「だ」([助动词特殊・ダ])的词。

例:「的」「げ」「がち」...

* 原来的 IPA 体系中,写作「名词接尾(形容动词词干)」。

5.1.27 名词-接尾-副词可能

与其他名词相连的结尾词中，可以起副词作用的词。

* 在 IPA 词性体系中，在上下文中实际起副词作用的词，后面附有〔副词的〕这一说明。在本体系中，无论具体用法如何，能作副词使用的词一律归入此类。

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

5.1.28 名词-接尾-助数词

与数词相连接组成名词的接尾词。比一般的「助数词」范围更广，包括接数词的普通名词。

例：「個」「つ」「本」「冊」「パーセント」「c m」「k g」「カ月」「か国」「区画」「時間」「時半」...

* 在 IPA 词性体系中，这些词中凡作为副词使用的都有「名词接尾助数词副词的」这一标记，但由于这是与用法相关的标记，本体系不包括。

5.1.29 名词-接尾-特殊

作为主要接用言的特殊接尾词，给予了重新定义。

例：「(楽し) さ」「(考え) 方」...

* 在 IPA 词性体系中，归到「名词接尾词」这类中。

5.1.30 名词-接续词性

连接单词与单词，起接续词作用的词。

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

5.1.31 名词-动词附属性

接在〔助词-接续助词〕的「て」后，有动词意义的词。

例：「ごらん」「ご覧」「御覧」「頂戴」...

注： 在原来的 IPA 词性体系中，有“名词引用字符串”这一标记，用来表示与单词不可分割的词，以及谚语、汉诗、方言、英语等。另外，有“名词公式”这一标记用来表示算术公式。由于很难将他们视为词性，因此本体系不认可它们是正式的词性名。现在，“名词引用字符串”中仅存「いわく」一词。

5.2 接头词

5.2.1 接头词-名词接续

后接名词（包括形容动词词干）的接头词，接数词者除外。

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

5.2.2 接头词-数词接续

后接名词的接头词中，与数词相接的词。

例：「約」「およそ」「毎時」など

5.2.3 接头词-动词接续

后接动词命令形或〔动词连用形〕+「なる／なさる／くださる」的接头词。

例：「お（読みなさい）」「お（座り）」...

5.2.4 接头词-形容词接续

后接形容词的接头词。

例: 「お（寒いですねえ）」「バカ（でかい）」...

5.3 动词

有关活用形的注意事项

未然形 在 THiMCO97 中，未然形细分为以下几类，但只要词尾无变化，都统称「未然形」。

- 未然レル接续
与ー（ラ）レル，ー（サ）セル相接的未然形。
例: 「読ま」「さ」...
- 未然ナイ接续
与ーナイ相接的未然形。
例: 「読ま」「し」...
- 未然ヌ接续
与ーヌ，ー（サ）シメル相接的未然形。
例: 「読ま」「せ」「来」...
- 未然ウ接续
与ー（ヨ）ウ相接的未然形。
例: 「読も」「し」...
& 在 ipadic1.0 以后，助动词ウ作为接续词。例如，对于「する」来说，「しよ」是其未然ウ接续。

连用形 除特殊的词尾以外，一律统称「连用形」。

- 连用マス接续
与ーマス相接的连用形。
例: 「読み」「し」「なさい」...
- 连用タイ接续
与ータイ，ーソウ，ーヅライ，一方（かた），逗号等相接的连用形。
例: 「読み」「し」「なさり」「向かひ」「習ひ」...
- 连用タ接续
ータ，ーテ相接的连用形。
例: 「読ん」「書い」「行っ」「問う」...

基本形 在 THiMCO97 中称为「词条形」的词。

与句号，体言，ーマイ等相接。
例: 「読む」「なさる」「問う」...

假定形 在 THiMCO97 中称为「假定バ接续」。

与ーバ，ードモ相接。
例: 「読め」「すれ」...

命令 i

力变・五段ラ行特殊命令形以及サ变・スルの命令形「せよ」的口語形。
例: 「来い」「なさい」「せい」...

命令 e

五段的命令形，文言已然形，一段动词的词干结句命令用法（仅「くれ」）。

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」是「(やめて) くれろ」的「ろ」缺省后的形态。「くれる」应看作一段动词中的特殊活用型。另外,「(やめて)(お) くれ(なさい)」的「くれ」是[动词-附属一段连用タイ接续],是不同的词,它的口語形态是「おくんない」。

命令 y o

一段・サ变・文言(力变)的命令形,以「y o」结尾。

例:「せよ」「みよ」「来よ」...

命令 r o

一段・サ变的命令形,以「r o」结尾。

例:「しろ」「みろ」...

ベキ接续

接在「ベキ」前,只有サ变这一情况。

例:「す」...

假定缩略 1

假定バ接续与「バ」相连,缩略后的形态<口語>。

例:「分かれりゃ」...

体言接续

仅出现在文言中。有不同于词条形的形态的词。

例:「助くる」(cf.「助く」)

体言特殊接续

以「る」结尾的动词接在「の」等后面时,发生音变的形态<口語>。

例:「(何) すん(の?)」...

体言接续特殊 2

「来る」「する」「とる」等「体言特殊接续」的词尾的「ん」缺省的形态。<口語>。

动词的活用类型一览(现代语)

在【活用形】的位置上加入下列词(没有【】这一标点)

5.3.1 动词-独立 力变 【活用形】

例:「くる」「来る」「やってくる」「やって来る」...

5.3.2 动词-附属 力变 【活用形】

例:「(て) くる」「(て) 来る」...

5.3.3 动词-独立 サ变・スル 【活用形】

「する」以及与[名词サ变接续]相接的「する」。

例:「する」...

5.3.4 动词-独立 サ変・スル 【活用形】

和语系的サ变动词。

例:「接する」...

* 仅将「ーし+ない」「ーせ+られる」「ーせ+ぬ」「ーし+よう」「ーする」「ーすれ+ば」「ーせよ」「ーしろ」的活用形视为[动词サ变・スル]。「ーし+」、「ーし+た」「ーし+たい」等的连用形一律视为[动词五段・サ行]。

5.3.5 动词-独立 ザ変・ズル 【活用形】

和语系的ザ变动词。

例:「信ずる」...

* 仅将「ーぜ+られる」「ーぜ+ぬ」「ーずる」「ーずれ+ば」「ーぜよ」「ーず+べし」的活用形视为[动词サ变・ズル]。「ーじ+ない」「ーじ+よう」的未然形和「ーじ+」、「ーじ+た」「ーじ+たい」等的连用形,以及「ーじろ」的命令形均视为[动词一段]。

5.3.6 动词-独立 一段 【活用形】

即上一段活用和下一段活用。

例:「着る」...

* 「病める」仅有[基本形]。

5.3.7 动词-附属 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」「きれ」「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)」的基本形是「いける」。

* 「(勉強)できる」不作为[附属]。

* 「うる」只有「うる」([基本形])和「うれ」([假定形])两种活用形。「う」是[动词文言基本形]。

5.3.8 动词-接尾 一段 【活用形】

即学校语法称为助动词的词。

例:「させる」「せる」「しめる」「しむる」「られる」「れる」...

5.3.9 动词-独立 五段・力行イ音变 【活用形】

与[助词接续助词]的「て」相连时发生イ音变的五段力行动词。

例:「解く」「聞く」...

5.3.10 动词-附属 五段・力行イ音变 【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 动词-附属 五段・力行促音变 【活用形】

与〔助词接续助词〕的「て」相连时发生促音变的五段力行动词。

例: 「いく」「行く」「ゆく」...

* 「ゆく」虽没有「ゆっ(て)」的形式, 但标为此类。「ゆき(て)」是〔动词文言连用タ接续〕。

5.3.12 动词-附属 五段・力行促音变 【活用形】

例: 「いく」「行く」「ゆく」「く<口語>」...

* 「ゆく」虽没有「ゆっ(て)」的形式, 但注此标记。「ゆき(て)」是〔动词文言连用タ接续〕。

5.3.13 动词-独立 五段・ガ行 【活用形】

与〔助词接续助词〕的「て」相连时发生イ音变的五段ガ行动词。

例: 「繼ぐ」「急ぐ」...

5.3.14 动词-独立 五段・サ行 【活用形】

与〔助词接续助词〕の「て」相连时不发生音变的五段サ行动词。

例: 「話す」...

5.3.15 动词-附属 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 动词-独立 五段・タ行 【活用形】

与〔助词接续助词〕的「て」相连时发生促音变的五段タ行动词

例: 「持つ」...

5.3.17 动词-独立 五段・ナ行 【活用形】

与〔助词接续助词〕的「て」相连时发生拨音变的五段ナ行动词

例: 「死ぬ」...

5.3.18 动词-独立 五段・バ行 【活用形】

与〔助词接续助词〕的「て」相连时发生拨音变的五段バ行动词

例: 「呼ぶ」...

5.3.19 动词-独立 五段・マ行 【活用形】

与〔助词接续助词〕的「て」相连时发生拨音变的五段マ行动词

例: 「進む」...

5.3.20 动词-附属 五段・マ行 【活用形】

例: 「こむ」「込む」...

5.3.21 动词-独立 五段・ラ行 【活用形】

与〔助词接续助词〕的「て」相连时发生促音变的五段ラ行动词

例: 「切る」「なる」...

5.3.22 动词-附属 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>」「じやる<口語>」「ぢやる<口語>」……

* 「なら(ない)」の基本形是「なる」。

5.3.23 动词-接尾 五段・ラ行 【活用形】

例: 「がる」…

5.3.24 动词-独立 五段・ラ行特殊 【活用形】

与助动词「ます」相连的形以及命令形为「ーい」的五段ラ行动词。

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 动词-附属 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 动词-独立 五段・ワ行ウ音变 【活用形】

与[助词接续助词]的「て」相连时发生ウ音变的五段ワ行动词

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

* [动词五段・ワ行促音变]的动词, 仅在与「て」相连的活用词尾为「う」的时候, 归入此类。除此以外, 视为[动词五段・ワ行促音变]。(人工修正数据中出现的是「ゆう」「食う」「すう」「負う」)

5.3.27 动词-附属 五段・ワ行ウ音变 【活用形】

例: 「たもう」「給う」…

5.3.28 动词-独立 五段・ワ行促音变 【活用形】

与[助词接续助词]的「て」相连时发生促音变的五段ワ行动词。

例: 「言う」「ゆう」「食う」「負う」「憂う」…

* 「憂う」并没有「憂って」, 但仍归入此类。(人工修正数据中出现的仅有「憂い(,)」这一形态。)

* [动词五段・ワ行促音变]的动词, 仅在与「て」相接的活用词尾是「う」的时候, 归入[动词五段・ワ行ウ音变]。

5.3.29 动词-附属 五段・ワ行促音变 【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>」…

动词的活用类型一览 (文言)

在IPA词性体系中, 并未对文言的活用类型进行详细分类。在IPADIC2.4以后, 虽然定义了四段, 上二段, 下二段等各种活用类型, 但辞典中的示例却并不充足。活用体系包括文言残留下来的词和用历史假名使用表示的口语。

5.3.30 动词-独立 四段・ハ行【活用形】

例: 「いふ」「云ふ」「向かふ」「習ふ」「思ふ」「能ふ」...

* 四段中, 还定义有「カ行」「ガ行」「サ行」「タ行」「バ行」「マ行」「ラ行」。

5.3.31 动词-独立 ラ変【活用形】

例: 「あり」「なり」「しかり」...

5.3.32 动词-独立 上二段・ハ行【活用形】

* 此外, 还有「ダ行」。

5.3.33 动词-独立 下二段・ア行【活用形】

* 此外, 还有「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」。

5.3.34 动词-独立 一段・得ル【活用形】

古文「得る」的活用类型。只有基本形和假定形。

5.4 形容词

除了把“词条形”“假定バ接续”“文言词条形”分别称为“基本形”“假定形”“文言基本形”之外, 本体系基本沿用了 THiMCO97 中所使用的活用形名称。另外, 把形容词的活用类型分为“形容词・アウオ段”“形容词・イ段”“形容词・文言”这几类。

未然又接续

接在一又前面的活用形。

例: 「寒から」...

未然ウ接续

接在一ウ前面的活用形。

例: 「寒かる」...

连用タ接续

接在一タ前面的活用形。

例: 「寒かつ」...

连用テ接续

接在一テ, 一ナイ, 一ナル, 一スル, 逗号等前面的活用形。

例: 「寒く」...

连用ゴザイ接续

接在一ゴザイマス前面的活用形。

例: 「寒う」「大きい」「のう」...

基本形

接在句号, 体言等前面。

例: 「寒い」「大きい」「ない」...

体言接续

文言活用中接在体言前的活用形。

例:「寒き」「なき」...

& 把「ーイ」算在基本形中。

假定形

接在ーバ前面的活用形。

例:「寒けれ」「なけれ」...

& 在 THiMCO97 中, 称作「假定バ接续」。

命令

文言活用中的命令形。

例:「よかれ」「美しかれ」...

& 把「ーイ」算在终止形中。

文言基本形

以ーシ结尾的活用形。

例:「良し」「遠し」「やむなし」...

假定缩略 1

假定バ接续与「バ」连用后缩略的活用形 1 <口语>.

例:「欲しけりゃ」「(それが) なけりゃ(困る)」...

假定缩略 2

假定バ接续与「バ」连用后缩略的活用形 2 <口语>.

例:「(それが) なきゃ(困る)」...

ガル接续

接在ーガル, ーゲ, ーソウ等前面的活用形。

例:「寒」「悲し」...

【活用形】的位置上加入后面的词。(没有【】这一标点)

5.4.1 形容词-独立 形容词・アウオ段 【活用形】

在形容词的活用类型中, 词干最后的母音以アウオ其中之一结尾的词。

例:「青い」「赤い」「厚い」「暑い」「熱い」...

* IPA 词性体系把「ない」的文言词条形「なし」定义为形容词的文言型活用的词条形。而本体系则把它定义「形容词・アウオ段(或形容词・イ段)」类型的“文言基本形”这一活用形。同样地, IPA 体系把「悪しき」定义为形容词的文言型活用的体言接续等, 本体系则把它和其他形容词一同定义为“体言接续”。

5.4.2 形容词-独立 形容词・イ段 【活用形】

形容词的活用类型中，词干最后的母音以イ结尾的词。

例: 「哀しい」「楽しい」「頼もしい」...

5.4.3 形容词-独立 形容词・イイ段 【活用形】

例: 「いい」「ええ」...

5.4.4 形容词-独立 形容词・不变化型 【活用形】

形容词中只有「基本形」这一用法的词。

例: 「カッコいい」...

5.4.5 形容词-附属 形容词・アウオ段 【活用形】

接在动词[连用タイ接续]或[连用タ接续]后面的形容词。

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

5.4.6 形容词-附属 形容词・イ段 【活用形】

接在动词[连用タイ接续]或[连用タ接续]后面的形容词。

例: 「らしい」「(て) ほしい」「(て) 欲しい」...

5.4.7 形容词-附属 形容词・イイ段 【活用形】

例: 「いい」...

5.4.8 形容词-附属 形容词・不变化型 【活用形】

接在动词[连用タイ接续]或[连用タ接续]后，只有「基本形」这一用法的形容词。

5.4.9 形容词-接尾 形容词・アウオ段 【活用形】

学校语法视为助动词的词。

例: 「(食べ) たい」...

5.4.10 形容词-接尾 形容词・イ段 【活用形】

学校语法视为助动词的词。

例: 「(嫌味) たらしい」...

5.5 副词

5.5.1 副词-一般

句子在它后面一定结束的词。不能作连体修饰的词。

例: 「あいかわらず」「多分」等

5.5.2 副词-助词类接续

后面可以接「の」「は」「に」「な」「する」「だ」等的副词。

例: 「こんなに」「そんなに」「あんなに」「なにか」「なんでも」...

5.6 连体词

5.6.1 连体词

只有修饰名词这一种存在形式的词。

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」「(ーも) 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 接续词

5.7.1 接续词

独立出现的接续词。

例: 「が」「けれども」「そして」「じゃあ」「それどころか」...

5.8 助词

5.8.1 助词-格助词-一般

即格助词。

* 「にて」也属于格助词。「の」既可作为格助词用, 也可用于连接名词和名词, 即所谓「AのB」这一用法, 后者称为「助词连体化」。

例: 「から」「が」「で」「と」「に」「へ」「より」「を」「の」「にて」...

5.8.2 助词-格助词-引用

紧跟在名词、人物的台词、会议决定、原因、判决、推测等后面的「と」。

例: 「(～だ) と(述べた。)」 「(～である) と(して執行猶予...)」...

5.8.3 助词-格助词-连词

连接格助词和动词, 主要起相当于格助词作用的词。

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」「に對し」「にたいして」「に對して」「にたいする」「に對する」「について」「につき」「につけ」「につけて」「につれ」「につれて」「にとって」「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」「をめぐって」「をめぐり」「をめぐる」「って<口语>」「ちゅう<関西弁>」「という>」「(何) ていう(人)<口语>」「っていう<口语>」「といふ」「とかいふ」.....

5.8.4 助词-接续助词

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) <口语>」「(行っ) ちゃ(いけない) <口语>」「(言っ) たって(しかたがない) <口语>」「(それがなく) ったって(平気) <口语>」.....

5.8.5 助词-系助词

例: 「こそ」「さえ」「しか」「すら」「は」「も」「ぞ」...

5.8.6 助词-副助词

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) <口语>」「(それ) じゃあ(よくない) <口语>」「ずつ」「(私) なぞ」「など」「(私) なり(に)」「(先生) なんか(大嫌い) <口语>」「(私) なんぞ」「(先生) なんて(大嫌い) <口语>」「のみ」「だけ」「(私) だって<口语>」「だに」「(彼) ったら<口语>」「(お茶) でも(いかが)」「等(とう)」「(今後) とも」「ばかり」「ばっか<口语>」「ばっかり<口语>」「ほど」「程」「まで」「迄」「(誰) も(が)(在[助词-格助词] 以及[助词-系助词] 前面的「も」)

5.8.7 助词-并列助词

例: 「と」「たり」「だの」「だり」「とか」「なり」「や」「やら」...

5.8.8 助词-终助词

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

* 在终助词「や」的使用中, 像「(まあいい) や」「(すごい) や」这些在关西方言中表示断定的「や」, 视为无活用变化的助动词。

5.8.9 助词-副助词／并列助词／终助词

在「か」一词中, 无法判断是副助词还是并列助词亦或终助词的, 例如, 下面的(a)(b)(c)中:

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

(b) 副词句中. 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」

* 最新的 IPA 词性体系将它再细分为「副助词」「并列助词」「终助词」这几类, 本体系不分。

5.8.10 助词-连体化

接在[名词]后面, 修饰体言的「の」。

* 在 THiMCO97 中, 将此用法的「の」归类为格助词。

5.8.11 助词-副词化

紧跟在拟音词、拟声词、拟态词及与此同类的名词、副词后面的「に」「と」。

例: 「に」「と」...

* 但是, 当修饰「する」「なる」表示状态变化时, 视为格助词。

5.8.12 助词-特殊

不能归入以上分类中的词。包括短歌和俳句等中所使用的助词等。

例: 「かな」「けむ」「(～しただろう) に」「(あんた) にゃ(わからん)」「(俺) ん(家)」...

5.8.13 助词-感动助词

在 IPA 辞典中原本没有的词。属文言性用法。ipadic2.4 以后给予了定义。

例: 「(松島) や」...

5.9 助动词

5.9.1 助动词 五段・ラ行アル 【活用形】

动词类型活用的助动词「である」「ではある」等的「ある」。

例: 「ある」...

5.9.2 助动词 五段・ラ行ゴザル 【活用形】

动词类型活用的助动词「ござる」。

例: 「ござる」...

5.9.3 助动词 形容词・イ段 【活用形】

形容词类型活用的助动词。

例: 「らしい」...

5.9.4 助动词 特殊・ナイ 【活用形】

否定助动词「ない」的活用类型。

例: 「ない」...

5.9.5 助动词 特殊・タ 【活用形】

表示终了的助动词「た」の活用类型。

例: 「た」「だ」...

* 由于「(学ん) だ」「(泳い) だ」这样接在五段活用的ガ行，ナ行，バ行，マ行后面时，表现形式变为「だ」，因此本体系把它们定义为分别有不同词干的语素。

5.9.6 助动词 特殊・ダ 【活用形】

断定助动词「だ」的活用类型。

例: 「だ」...

5.9.7 助动词 特殊・デス 【活用形】

断定助动词「です」的活用类型。

例: 「です」...

5.9.8 助动词 特殊・ドス 【活用形】

断定助动词「どす」的活用类型。

例: 「どす」...

5.9.9 助动词 特殊・ジャ 【活用形】

断定助动词「じゃ」的活用类型。

例: 「じゃ」...

* 表断定的「だ」浊化后的词。

5.9.10 助动词 特殊・マス【活用形】

表谦逊 自谦的助动词「ます」的活用类型。

例: 「ます」...

5.9.11 助动词 特殊・ヌ【活用形】

否定助动词「ぬ」的活用类型。

例: 「ぬ」...

5.9.12 助动词 特殊・ヤ【活用形】

在关西方言中表断定的助动词。

例: 「(そう) や」...

5.9.13 助动词 不变化型【活用形】

在现代语中没有活用的助动词。包括无法确定活用形式的口语和方言。

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

5.9.14 助动词 文言・??【活用形】

文言助动词。目前给出了定义的活用类型如下: 「文言・ベシ」「文言・ゴトシ」「文言・ナリ」「文言・マジ」「文言・シム」「文言・キ」「文言・ケリ」「文言・ル」「文言・リ」

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

* IPA 词性体系中有「じ」的活用类型, 但实际上它是不变化的, 因此本体系将它视为不变化型。

5.10 感动词

5.10.1 感动词

感动词, 寒暄。

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 标点

5.11.1 标点-普通

除以下分类以外的, 一般性的标点。

例: 「○」「◎」「@」「\$」「〒」「→」「+」等

5.11.2 标点-字母

英语字母, 包括大写字母和小写字母。

例: 「A」「a」...

5.11.3 标点-句号

即句号。

例: 「.」「.」...

5.11.4 标点-逗号

即逗号。

例: 「,」「,」「,」...

5.11.5 标点-空格

全角的空格字符（不可视）。

5.11.6 标点-左括号

例: 「(」「{」「[」「[「」「[「」「[「」「[「」...

5.11.7 标点-右括号

例: 「)」「}」「]」「]」」「]」」「]」」「]」」...

5.12 补白

5.12.1 补白

口语中附和他人说话时的词，以及插入性的声音。

例: 「あの」「うんと」「えと」...

5.13 其他

5.13.1 其他-感动

无法归为[名词-接尾词]及[助词-终助词]的词。

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連接規則の実装」, 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『茶筌』」, 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付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.

附录

A 变更点

A.1 从 ipadic-2.6.3 到 ipadic-2.7.0 的变更点

变更点记录在 <http://chasen.aist-nara.ac.jp/%7emasayu-a/ipadic/arch/ipadic-2.7.0-diff.txt>

- 旭化成の水嶋先生报告了很多未登录词和错误事例。
- 使用另外的文件 Onebyte.dic 保存单字节(1byte)词语。
- 更改活用形和活用类型的定义，增加了「形容词・イイ」和「特殊・ドス」。

A.2 从 ipadic-2.6.2 到 ipadic-2.6.3 的变更点

变更点记录在 <http://chasen.aist-nara.ac.jp/%7emasayu-a/ipadic/arch/ipadic-2.6.3-diff.txt>

- 国语研的峰布由纪报告了错误事例。
- 顺天大学的宋浩报告了错误事例。

A.3 从 ipadic-2.6.1 到 ipadic-2.6.2 的变更点

改进了连接表参数。辞典词目没有变化。

A.4 从 ipadic-2.6.0 到 ipadic-2.6.1 的变更点

辞典词目没有变化。

- 修正了 chasenrc
- 与 chasen2.3.2 的对应

A.5 从 ipadic-2.5.1 到 ipadic-2.6.0 的变更点

ipadic-2.5.1 与 ipadic-2.6.0 采用完全相同的词性体系。

- 收录了長崎大学的鹤丸先生提供的未登录词（约 6000 词），在此表示感谢。
变更点在 <http://chasen.aist-nara.ac.jp/%7emasayu-a/ipadic/arch/ipadic-2.6.0-diff.txt>。
- 增加的其他单词在 <http://chasen.aist-nara.ac.jp/%7emasayu-a/ipadic/arch/ipadic-2.6.0-add.txt>。
- 与 chasen2.3.1 的对应
 - 删除了.pat, .ary 生成部分
 - 与双排序辞典 dadic 的对应
 - 与文字代码变更选项-i 的对应

A.6 从 ipadic-2.5.0 到 ipadic-2.5.1 的变更点

ipadic-2.5.0 与 ipadic-2.5.1 采用完全相同的词性体系。

- 与「助词-格助词-连語」相当的词句建立单独的文件 Postp-col.dic。
- 对一段动词和助动词周边的参数进行了变更。
- 新收录词语列表在 <http://chasen.aist-nara.ac.jp/%7emasayu-a/ipadic/arch/ipadic-2.5.1-newword.txt>。

A.7 从 ipadic-2.4.X 到 ipadic-2.5.0 的变更点

ipadic-2.4.X 与 ipadic-2.5.0 采用完全相同的词性体系。

B 辞典的版权

本辞典包含了 ICOT 的研究成果。

进行二次发布时，请附上以下文字：

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.