Handle with care(取扱注意)

TR-I-0272

日本語形態素解析関連プログラム
ユーザーズ・マニュアル

Users' Manual for Japanese Morphological Analysis and Related Programs

浦谷 則好
Noriyoshi URATANI

1992. 8

概要

本レポートは、日本語のテキストを形態素解析するプログラムとそれに関連するプログラムの操作方法を説明しています。

ＡＴＲ自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

目次

第１章

日本語形態素解析プログラム（ＡＮＡ）
ユーザーズ・マニュアル
ＡＮＡ　Users'　Manual

この章では、日本語のテキストを形態素解析するプログラム（ＡＮＡ）と結果を表示するためのプログラム（ＶＭ）の操作方法を説明します。

ソフトウェアの概要

名　称　：	ANA(日本語形態素解析プログラム)
用　途　：	日本語テキストの形態素解析
機　能　：	日本語テキストを辞書情報に基づいて、形態素解析するプログラムである。対話モードとバッチモードがあり、対話モードはメニューの選択で辞書やテキストを指定するようになっている。
使用言語：	C および VAX LISP
使用環境：	OS： UNIX (DEC ULTRIX) 主記憶： 8MB以上 ディスク容量： 6MB以上 利用可能機種： VAX その他の制限： 対話モードはCIT-600 端末用に設計してある。
ソフトウェアの規模：	約 5MB
提供の形態：	8mm データ カートリッジ
文書：	日本語マニュアル
備考：	対話モードは CIT-600 端末用なので、その他の端末での動作は保証しない。関連ソフトとして結果を表示するためのプログラム (VM)を付属してある。
担当者：	浦谷　則好

日本語形態素解析プログラム

日本語テキストファイルを読み、形態素情報データを自動的に作成するプログラムである。

Usage:

```
ａｎａ                                          ・・・・対話モード
ａｎａ　辞書名　分野名　テキストファイル．．．　　・・・・バッチモード
```

引数の説明

辞書名　　解析に利用する辞書の指定
辞書はディレクトリ $JMADICT／SDICの下に
なければならない

分野名　　解析結果の出力先
ディレクトリ $JMADATA／ａｎａ　の下のディレクトリ
を指定する

テキストファイル
解析を行うファイル名（複数可）を指定する

（例）

```
ｃｄ　／data2／CONV／TELEPHONE
ａｎａ　SD_ALL　ｔｅｌ　ｔｅｌ－data1　ｔｅｌ－data2
```

辞書 SD_ALL を使用して、ｔｅｌ－data1, ｔｅｌ－data2 を解析する。
結果は分野で指定したディレクトリ（$JMADATA／ａｎａ／ｔｅｌ）の下に入る。
出力ファイル名は　WO_ｔｅｌ－data1－01 および WO_ｔｅｌ－data2－01
等となる。

環境

環境変数JMAHOME が指定されていて、環境変数JMADICT が$JMAHOME／ｄｉｃｔ
に指定され、その下のディレクトリSDIC の下に辞書ファイルが存在している必要
があります。また、ディレクトリ$JMAHOMEの下にはディレクトリｅｘｅｃ，ｐｕｂ
が存在し、それぞれの下に必要なファイル（ａｎａ４．ｓｕｓおよびｈｉｎｓｈｉ等）が
存在しなければなりません。さらに、JMADATAが指定されていて、その下にディ
レクトリ ａｎａが存在し、そのａｎａの下に分野で指定するディレクトリが存在して
いる必要があります。

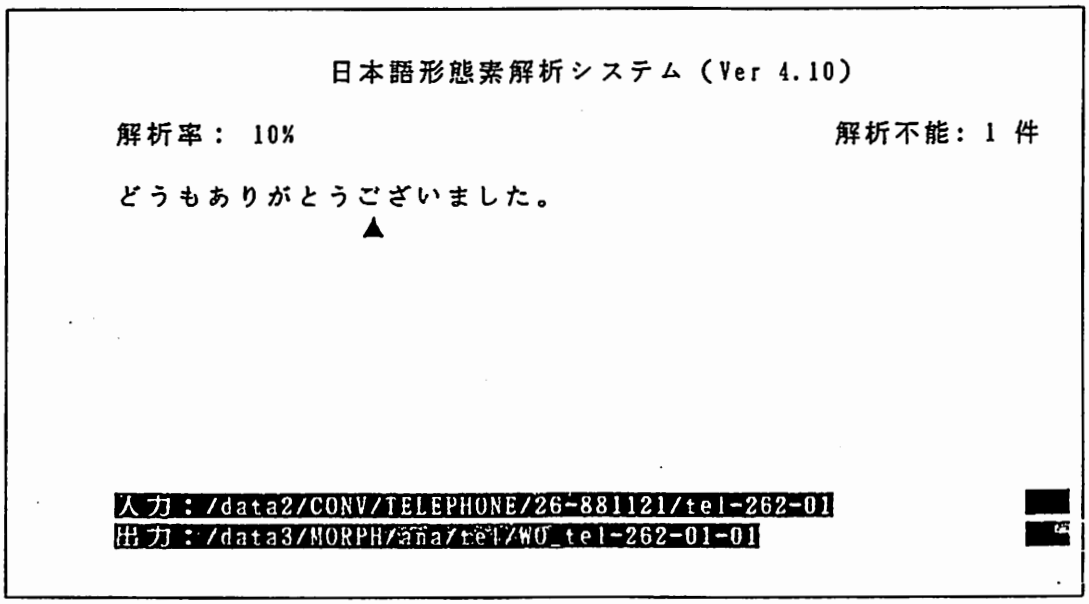
入力ファイルの例

通訳者：もしもし、国際コンピューター会議の事務局でしょうか。
事務局：はい、こちら国際コンピューター会議の事務局ですが。
通訳者：私、トロント大学のクリス・マーローと申します。
（たち）確かそちらで講演をしておりますジェニファー・ホスキンス先生とお話できないかと思いまして、お電話かけております。
事務局：［えー］いま先生はちょうど講演中でございまして、［えー］講演が終わりましたら連絡を取るようにいたしますが、［えー］何か（と）緊急な御用事でしょうか。
通訳者：［えー］いささか緊急の用事ですので、［えー］先生はどのぐらい、あと講演にかかりますでしょうか。
通訳者：いま日本は何時ですか。
事務局：いま三時十分です。
通訳者：いますぐにでもお話したいところですが、もし無理ならば、先生の御講演の直後にお話したいんですが、お話できますでしょうか。
事務局：［えー］それは問題ないと思います。
［えー］私どもの事務局の人間が、先生の講演終わりましたら、この事務局までお連れしたいと思いますが、こちらからお電話していただきましょうか。
それとも［お］四時、そうですね、四時十分ぐらいにお電話を再度いただけますでしょうか。
通訳者：私のオフィスの方にお願いします。
至急ということではありますが、緊急事態ということではありませんので、［え］講演の後、オフィスの方にお電話下さるようお伝え下さい。
事務局：わかりました。
それでは再度恐れ入りますが、お名前と、念のため電話番号を教えていただけますでしょうか。
通訳者：はい、私の名前はクリス・マーローと申します。
大学は先生もよく御存知ですが、電話番号は二丸六、四五一、八丸四三です。
事務局：［えー］それでは再度確認します。
［え］電話番号が二零六の四五一、八零四三ですね。
通訳者：そのとおりです。
ではお電話お待ちしております。
どうもいろいろありがとうございました。
事務局：どうもありがとうございました。
それでは先生にお電話していただくようにしますので。

対話モードでの操作方法

１．	ana [ret]	システムの起動
２．	CL>(run) 実行	
３．	解析辞書を選択して下さい：	解析辞書と選択番号が表示されているので、一つ選択する。
４．	分野名を選択して下さい：	分野と選択番号が表示されているので、一つ選択する。
５．	日本語テキストのディレクトリ名を入力して下さい：	テキストファイルのディレクトリを入力する。リターンだけの場合は解析処理を開始する（→７）。
６．	テキストファイルの選択	ディレクトリを入力すると、その下にあるファイルを選択番号を付けて表示する。解析を行なうファイルの選択番号を入力する。複数のファイルを選択する時は、番号を複数個入力する。入力後、ディレクトリの入力へ戻る（→５）。
７．	解析の開始	指定したファイル全てについて、解析処理を行なう。

解析画面は左上にテキストファイルの解析率、右上に解析不能文の件数、それから解析中の文とその下の三角は解析処理をしている語の位置、画面下には入出力ファイル名が表示されている。



日本語形態素表示プログラム

日本語形態素解析結果のファイルを表示するためのプログラムである。

Usage:

ｖｍ －[Ａａｌｄｓｔｎ]　形態素解析結果ファイル．．．

引数の説明

オプション

A	ファイル中のすべての情報を出力
a	ファイル中のすべての情報を出力（Aと同じ）するが、
.	長い語は先頭７文字にカットする
l	語を長形式で表示
d	スラッシュで語を区切る
s	空白で語を区切る
t	形態素の合計数を出力
n	文番号の表示を行う

形態素解析結果ファイル

形態素解析結果のファイル（複数可）を指定する

出力例

出力ファイルの例を後に揚げる

ｖｍの出力ファイルの例

```
##################
### Option: A
##################
FILE : ED_ifg-012-10

質問者：                                        記　号      －－      －－

もしもし        もしもし        もしもし        感動詞      －－      －－
。                                              記　号      －－      －－

事務局：                                        記　号      －－      －－

はい            はい            はい            感動詞      －－      －－
。                                              記　号      －－      －－

会議            かいぎ          会議            普名詞      －－      －－
事務            じむ            事務            普名詞      －－      －－
局              きょく          局              普名詞      －－      －－
で              で              だ              助動詞      －－      連用
ございます      ございます      ございます      補動詞      特殊      終止
。                                              記　号      －－      －－

質問者：                                        記　号      －－      －－

会議            かいぎ          会議            普名詞      －－      －－
の              の              の              格助詞      －－      －－
宿泊            しゅくはく      宿泊            普名詞      －－      －－
施設            しせつ          施設            普名詞      －－      －－
について        について        について        格助詞      －－      －－
お              お              御              接頭語      －－      －－
尋ね            たずね          尋ねる          本動詞      下一      連用
し              し              する            補動詞      サ変      連用
たい            たい            たい            助動詞      －－      連体
の              の              の              準助詞      －－      －－
です            です            です            助動詞      －－      終止
が              が              が              接助詞      －－      －－
。                                              記　号      －－      －－

##################
### Option: a
##################
FILE : ED_ifg-012-10

質問者：                                        記　号      －－      －－

もしもし        もしもし        もしもし        感動詞      －－      －－
。                                              記　号      －－      －－
```

事務局：			記　号	－－	－－
はい	はい	はい	感動詞	－－	－－
。			記　号	－－	－－
会議	かいぎ	会議	普名詞	－－	－－
事務	じむ	事務	普名詞	－－	－－
局	きょく	局	普名詞	－－	－－
で	で	だ	助動詞	－－	連用
ございます	ございます	ございます	補動詞	特殊	終止
。			記　号	－－	－－
質問者：			記　号	－－	－－
会議	かいぎ	会議	普名詞	－－	－－
の	の	の	格助詞	－－	－－
宿泊	しゅくはく	宿泊	普名詞	－－	－－
施設	しせつ	施設	普名詞	－－	－－
について	について	について	格助詞	－－	－－
お	お	御	接頭語	－－	－－
尋ね	たずね	尋ねる	本動詞	下一	連用
し	し	する	補動詞	サ変	連用
たい	たい	たい	助動詞	－－	連体
の	の	の	準助詞	－－	－－
です	です	です	助動詞	－－	終止
が	が	が	接助詞	－－	－－
。			記　号	－－	－－

```
####################
### Option: y
####################
FILE : ED_ifg-012-10
```

(質問者：)
もしもし(。)
(事務局：)
はい(。)
かいぎじむきょくでございます(。)
(質問者：)
かいぎのしゅくはくしせつについておたずねしたいのですが(。)

```
####################
### Option: l
####################
FILE : ED_ifg-012-10
```

質問者：もしもし。
事務局：はい。
会議事務局でございます。
質問者：会議の宿泊施設についてお尋ねしたいのですが。

```
##################
### Option: ld
##################
FILE : ED_ifg-012-10


質問者：/ もしもし/ 。/
事務局：/ はい/ 。/
会議/ 事務/ 局/ で/ ございます/ 。/
質問者：/ 会議/ の/ 宿泊/ 施設/ について/ お/ 尋ね/ し/ たい/ の/ です/ が/ 。/

##################
### Option: ls
##################
FILE : ED_ifg-012-10


質問者：  もしもし  。
事務局：  はい  。
会議  事務  局  で  ございます  。
質問者：  会議  の  宿泊  施設  について  お  尋ね  し  たい  の  です  が  。

##################
### Option: t
##################
FILE : ED_ifg-012-10


質問者：
もしもし。
事務局：
はい。
会議事務局でございます。
質問者：
会議の宿泊施設についてお尋ねしたいのですが。

総単語数 : 26
##################
### Option: n
##################
FILE : ED_ifg-012-10


[1] 質問者：
[2] もしもし。
[3] 事務局：
[4] はい。
[5] 会議事務局でございます。
[6] 質問者：
[7] 会議の宿泊施設についてお尋ねしたいのですが。
```

第２章

日本語形態素検索プログラム（Retrieve）
ユーザーズ・マニュアル
Retrieve Users' Manual

この章では、日本語形態素解析結果ファイルの中身を検索するプログラム（Retrieve）の操作方法を説明します。

ソフトウェアの概要

名　称　：	Retrieve(日本語形態素検索プログラム)
用　途　：	日本語形態素解析結果のファイルの中身の検索
機　能　：	日本語形態素解析プログラム(ANA)の出力ファイル(日本語形態素解析結果)の中身を検索するプログラムであり、形態素数のカウント、品詞の出現頻度の算出、指定したパターンの出現の検索などが可能である。
使用言語：	C
使用環境：	OS： UNIX 主記憶： 4MB以上 ディスク容量： 500KB以上 利用可能機種： VAX ほか その他の制限： 特になし
ソフトウェアの規模：	約32 KB
提供の形態：	8mm データカートリッジ
文書：	日本語マニュアル
備考：	特になし
担当者：	浦谷　則好

日本語形態素検索プログラム

日本語形態素解析プログラム（ＡＮＡ）の結果ファイルの中身を検索するプログラムである。

1．　Usage:

retrieve [-w workfile] [-k] [-o outfile] morph-file... → 2
retrieve -w workfile [-k] [-o outfile] → 3
retrieve [-w workfile] [-k] [-o outfile] -f procfile morph-file... → 4
retrieve -w workfile [-k] [-o outfile] -f procfile → 5

引数の説明

-w workfile	検索用ワークファイル名の指定 省略時は /tmp に適当に作られる
-k	検索用ワークファイルの保存 省略時はプログラム終了時に消去
-o outfile	検索結果を入れるファイル名の指定 省略時はディスプレイに出力
-f procfile	検索コマンドを入れてあるファイル名の指定 省略時は対話モードになる
morph-file	形態素解析結果ファイル（複数指定可）

2．　形態素ファイルから検索用ワークファイルを作成し、検索するモード

ワークファイルを残したい場合は -w および -k オプションを指定する。

```
ex.    retrieve  -w  TEL.WRK  -k  ED_tel-001
```

3．　既存のワークファイルを使って検索するモード

ワークファイルは必ず指定しなければならない。実行後もワークファイルを残したければ -k オプションを指定する。

```
ex.    retrieve  -w  TEL.WRK  -k
```

4．　コマンドファイルを用いて検索するモード（既存のワークファイルが無い場合）

予めコマンドファイルにはエディタ等を用いて検索コマンドを入れておく。

```
ex.   TEL.cmd

        S 電話
        F
        @a=本動詞
        W  @$a
```

retrieve -f TEL.cmd ED_tel-001-*

ワークファイルを保存したい場合は -w と -k オプションを指定する
ex. retrieve -f TEL.cmd -w TEL.WRK -k ED_tel-001-*

検索結果をファイルに出力したい場合は -o オプションを指定する
ex. retrieve -f TEL.cmd -o TEL.OUT -k ED_tel-001-*

→ 6

5．既存のワークファイルを使ってコマンドファイルの内容で検索するモード

ワークファイルとコマンドファイルは必ず指定する。処理後も
ワークファイルを消したくなければ -k オプションを指定する
ex. retrieve -f TEL.cmd -w TEL.WRK -k

検索結果をファイルに出力したい場合は -o オプションを指定する
ex. retrieve -f TEL.cmd -w TEL.WRK -k -o TEL.out

→ 6

6． ./.retrc が 実行される

検索システムの初期設定コマンドファイル

./.retrc には通常、毎回 決まりきった処理を行うコマンドを予め入れておく

ex.

```
x
x          RETRIEVE SYSTEM
x
@v=本動詞 | 補助動詞
T
```

- この様な内容を入れておくと、起動時にメッセージを表示後
  形態素数を表示する
- そして検索時$vで本動詞、補助動詞を指すようになる

→ 7

7．コマンドの説明

コマンドの入力時はシステムからプロンプト(>) が表示される

コマンドは次の形式をとる

コマンド [パラメータ] [出力ファイル]

コマンドは１文字の英字で文字の大小はどちらでもよい。基本的には１行に１つのコマンドが入力されるものとする。１つ以上のコマンドを定義したファイルをコマンド定義ファイルと言い、起動時 -f オプションで指定が可能である。
コマンド行の後ろに " > 出力ファイル " を指定する事により検索結果をファイルに出力する事ができる。これは -o オプションで指定された出力ファイルより、優先され、ＵＮＩＸのリダイレクトと違い、この出力ファイルはアペンド形式で出力される。そのコマンドの出力だけ一時的に出力先を切り替える

ex. s 電話 > denwa [ret]

コマンドが１行で収まらない時はバックスラッシュ（＼）を使う事により複数行に継続する事が可能である。

```
ex.  w [ @ 記号¥    [ret]
     +  *  ¥        [ret]
     + ] @ 記号     [ret]
```

尚、対話モードでは ^D の入力により処理を終了する

コマンド　パラメーター

N　　n　　S,W コマンドにおいて検索結果を前後をｎ語で出力する
初期値は 0

S,W コマンド実行時、形態素情報以外に出現データを
文として出力するが、その時、該当形態素の前後ｎ語を
同時に出力し、前後の文のつながりを見るためのもの
である。

```
ex.> S です   [ret]
   です
   助動詞 —— 終止 —— です
   > N1

   >S  です
   そうです。
   助動詞 —— 終止 —— です
```

T　　-　　検索対象となっている形態素数の表示

```
ex.> T  [ret]
   1200
```

X　　任意の文字列　　コメントを出力する

```
ex.> X TEST  [ret]
   TEST
```

@　　代入文　　変数への代入
変数名は任意の英数字とし、扱えるデータは文字型と
する
" $ 変数名 " として参照が可能

```
ex.> @A=TEST          [ret]
   > X $A             [ret]
   TEST
   > @V=本動詞        [ret]
   > w @$V
   行く
   本動詞    五段    終止    行く
```

F　　-　　品詞の出現頻度の出力

```
ex.> F                [ret]
   100     記号      ----     ----
    15     形容詞    ----     未然
    16     形容詞    ----     連用
    12     形容詞    ----     終止
    14     形容詞    ----     連隊
                .
                .
                .
```

I　　n△範囲　　n語の連語情報の出力
n語の連語情報を集計し、頻度と形態素を出力する
範囲は出力する頻度の範囲でｎ１－ｎ２型式で
記述する

ex.　I 3　　[ret]　　3語の連語情報
I 3 5　　[ret]　　3語の連語情報の内頻度が5のもの
I 3 100- [ret]　　3語の連語情報の内頻度が100以上のもの
I 3 -10　[ret]　　3語の連語情報の内頻度が5のもの
I 3 100-200 [ret]　3語の連語情報の内頻度が100以上200以下

ex.>I 3 -10　[ret]

10	ですが、		
助動詞	----	終止	です
接助詞	----	----	が
記号	----	----	、

8	参加して		
サ名詞	----	----	参加
補動詞	サ変	連用	し
接助詞	----	----	て

P　　n△範囲　　n語の品詞の連体情報の出力
n語の品詞の連体情報を集計し、頻度を出力する
範囲は出力する頻度の範囲でn1-n2 型式で記述する

ex.>p 3　60　　[ret]

60		
記号	----	----
感動詞	----	----
記号	----	----

S　　文字式　　形態素中の任意の文字列を検索する
文字式にはワイルドカード(*)，文字選択([ ])
が指定可能である

S電話
Sでしょう

ワイルドカード（*）を使用する事が可能である。

Sお*する
S {*}

カギ括弧（[ ]）で囲む事により複数文字のOR条件が
記述可能である。

S [おご] * なさい

ex.　> S みたい
してみたいと思い

補動詞	変則	連用	み
助動詞	----	終止	たい

W　　一般式　　特定の形態素の検索
一般式を使い、様々な形態素の組み合わせを検索する。

一般式の書式

一般式は形態素の指定（語、品詞、活用型、活用形）を一つ
もしくは複数を ‘＋’ でつないだものとする。

一般式：＝ 形態素指定｜
　　一般式 ′＋′ 形態素指定

形態素の指定は語と品詞の指定を組み合わせたものとする。
語 もしくは品詞のどちらかが省略されている場合も許可する
品詞の指定の前には必ず ′@′ をつける。

形態素指定：＝ 語指定｜
　　語指定 ′@′ 品詞式｜
　　′@′ 品詞式

語の指定は文字式と同じである
先頭に ′R′ を付けると正規表現の検索を行う。

語指定：＝ 語｜
　　′R′語

品詞式には｜(NOT)、′｜′(OR)、 ′(…)′（括弧）が使用可能である。

式は左から右に評価され、演算子は括弧、ＮＯＴ、ＯＲの順で優先される。

品詞式：＝ 品詞指定｜
　　品詞式 ′｜′ 品詞指定｜
　　′(′ 品詞式′)′｜
　　′!′ 品詞式

品詞指定は必ず一つの品詞が指定されその後ろに′：′ で区切って活用型、
活用形が指定される。

品詞指定：＝ 品詞｜
　　品詞 ′：′ 活用型｜
　　品詞 ′：′ 活用形｜
　　品詞 ′：′ 活用型 ′：′ 活用形

一般式の例

接頭語「お］、「ご」と補助動詞にはさまれる語

　　「おご」@接頭辞+*+@補動詞

本動詞連用形と補助動詞

　　@本動詞：連用+@補動詞

接尾辞「時」と接尾辞「間」

　　時@接尾辞+間@ 接尾辞

形容名詞と助動詞 「です」

　　@形名詞+Rです@助動詞

接頭辞「お」と形容名詞もしくはサ変名詞

　　お@接頭辞+@形名詞｜サ変名詞

サ変名詞と補助動詞以外のもの

　　@サ変名詞+@!補動詞

「電話」と言う語を含む固有名詞

　　電話@固名詞

ex:

> w @ 補助動詞 + たい @ 助動詞

参加して頂きたいと思っ

補動詞　五段　連用　頂き
助動詞　----　終止　たい

してみたいと思い

補動詞　変則　連用　み
助動詞　----　終止　たい

第３章

日本語形態素辞書編集プログラム（Edkdic）
ユーザーズ・マニュアル
Edkdic Users' Manual

この章では、日本語のテキストを形態素解析するプログラム（ＡＮＡ）で用いる辞書の編集プログラム（Edkdic）、形態素解析結果ファイルから辞書情報を自動的に抽出するプログラム（Mo_extr）、および辞書構築プログラム（Sdicgen）の操作方法を説明します。

ソフトウェアの概要

名　称　：	Edkdic(日本語形態素辞書編集プログラム)
用　途　：	日本語形態素解析で用いる辞書の編集
機　能　：	日本語形態素解析プログラム(ANA)で使用する辞書の編集を行うプログラムであり、形態素抽出結果を辞書に反映させるバッチモードと対話的に追加・削除・修正を行う対話モードがある。
使用言語：	C および VAX LISP
使用環境：	OS： UNIX (DEC ULTRIX) 主記憶： 4MB以上 ディスク容量： 12MB以上 利用可能機種： VAX その他の制限： 対話モードは CLT-600端末用に設計してある。
ソフトウェアの規模：	約 7MB
提供の形態：	8mm データ カートリッジ
文書：	日本語マニュアル
備考：	対話モードは CIT-600 端末用なので、その他の端末での動作は保証しない。(動くものもあれば動かないものもある。) 関連ソフトとして 辞書情報を抽出するプログラム (Ma extr), 実際に辞書を構築するプログラム(Sdicgen) を付属してある。
担当者：	浦谷　則好

日本語形態素辞書編集プログラム

日本語形態素解析プログラム（ＡＮＡ）で用いる辞書の編集をするためのプログラムです。

Usage:

```
edkdic                                          ・・・・対話モード
edkdic   辞書名  分野名  テキストファイル        ・・・・バッチモード
```

引数の説明

辞書名　　$JMADICTの下にある辞書用のディレクトリを指定

分野名　　$JMADATA/freqの下にあるテキストファイルのある
ディレクトリを指定します

テキストファイル
辞書情報を抽出するためのファイル名を指定します。
ファイルはmo_extr（後述）の出力ファイル（FD_で始まっているもの）でなければなりません。

（例）

```
edkdic  work  talk  FD_ifg-012-09
```

環境

環境変数JMAHOMEが指定されていて、環境変数JMADICTが$JMAHOME／dictに指定されている必要があります。辞書名で指定するディレクトリの下にはディレクトリHIRA,KATA,KANJ,ETCとファイルIdiom,maxfreqが存在している必要があります。また、ディレクトリ$JMAHOMEの下にはディレクトリexec,pub,libが存在し、それぞれの下に必要なファイル（edkdic.susおよびhinshi,keyboard.fas等）が存在しなければなりません。さらに、JMADATAが指定されていて、その下にディレクトリfreqが存在している必要があります。

実行内容	画面出力	入力方法
1 実行命令	T CL>	(run) … 2 へ
2 メニュー出力	[1] マニュアルモード [2] AUTOモード [9] 処理終了 処理モード：	1 … 7 へ 2 … 3 へ 9 … 作業終了
3 辞書選択	[0] imp [1] jpn [2] key [3] new [4] ppr [5] tel 辞書を選択して下さい：	入力したい辞書を選択
4 ディレクトリ選択	[0] imp [1] jpn [2] key [3] new [4] ppr [5] tel ディレクトリを選択して下さい：	抽出ファイルのディレクトリを選択 … 5 へ
5 ファイル選択	[0] FD_ppr-001-01 [1] FD_ppr-001-02 [2] FD_ppr-001-03 [3] FD_ppr-001-04 [4] FD_ppr-001-05 [5] FD_ppr-001-06 [6] FD_ppr-001-07 ファイルを選択して下さい：	抽出ファイルの番号を選択 注1 … 6 へ
6 辞書に組込みたくない品詞を選択	(1) 形容詞 (2) 形容詞 (3) 普名詞 (5) サ名詞 (6) 代名詞 (7) 数 詞 (8) 副 詞 (9) 連体詞 (10)接続詞 (17)接尾語 (18)接尾語 (19)補動詞 (30)固名詞 (32)本動詞 辞書に組み込まない品詞を選択：	辞書に組み込みたくない品詞を選択する 注2 … 全部組み込まれる … 2 へ
7 辞書選択	(0) imp (1) jpn (2) key (3) new (4) ppr (5) tel 辞書を選択して下さい：	… 8 へ
8 正しい語の入力	語を入力して下さい：	正しい語を入力する 辞書にある … 9 へ 辞書にない … 13 へ … 2 へ

実行内容	画面出力	入力方法
9	(1) 漢字：聞く 読み：き・く 品詞：本動詞*五段 漢字2： 派生語： 頻度：(5) コマンドを入力して下さい＞ A … データの追加　B …前ページ D … データの削除　F …次ページ M … データの修正　X …終了	A … 10 へ B … 前のページに もどる D … 11 へ F … 次のページに 進む M … 12 へ X … 8 へ
10	(1) 漢字：聞く 読み：き・く 品詞：本動詞*五段 漢字2： 派生語： 頻度：(5) 語幹を入力して下さい＞ 読みを入力して下さい＞ 漢字を入力して下さい＞ 品詞を入力して下さい＞ SUB漢字を入力して下さい＞ 語幹活用形を入力して下さい＞ 確認〔Y/N〕＞ A … データの追加　B …前ページ D … データの削除　F …次ページ M … データの修正　X …終了	それぞれに入力する なければ 注3 Y … 9 へ N … 9 へ
11 辞書を削除する	(1) 漢字：聞く 読み：き・く 品詞：本動詞*五段 漢字2： 派生語： 頻度：(5) 番号を入力して下さい＞ A … データの追加　B …前ページ D … データの削除　F …次ページ M … データの修正　X …終了	削除したい番号を入力 ↓ 確認〔Y/N〕 Y … 9 へ N … 9 へ
12 辞書の修正	(1) 漢字：聞く 読み：き・く 品詞：本動詞*五段 漢字2： 派生語： 頻度：(5) 番号を入力して下さい＞ A … データの追加　B …前ページ D … データの削除　F …次ページ M … データの修正　X …終了	修正したい番号を入力 … 14 へ

実行内容	画面出力	入力方法
13 新たに辞書を作る	辞書がありません。 新たに作りますか(Y/N)：	Y … 15 へ N … 8 へ
14	〔旧〕 漢字：聞く 読み：き・く 品詞：本動詞*五段 漢字2： 派生語： 頻度：(5) 語幹を入力して下さい> [redacted] 読みを入力して下さい> [redacted] 漢字を入力して下さい> [redacted] 品詞を入力して下さい> [redacted] SUB漢字を入力して下さい> [redacted] 語幹活用形を入力して下さい> [redacted] 確認〔Y/N〕> [redacted]	それぞれに入力する なければ 注3 確認〔Y/N〕 Y … 9 へ N … 9 へ
15	語幹を入力して下さい> [redacted] 読みを入力して下さい> [redacted] 漢字を入力して下さい> [redacted] 品詞を入力して下さい> [redacted] SUB漢字を入力して下さい> [redacted] 語幹活用形を入力して下さい> [redacted] 確認〔Y/N〕> [redacted]	それぞれに入力する なければ 注3 確認〔Y/N〕 Y … 8 へ N … 8 へ

注1：ファイルが多く有るため１ページに入りきれない時、^N で下にスクロールする
注2：２ヶ以上選択する場合、ブランクでむすぶ。
注3：品詞がいくつも入力できるようになっている。
　　品詞の入力が終了したら空打ち( )をする。

日本語形態素抽出プログラム

日本語形態素解析結果のファイル（修正済みのもの）から辞書編集プログラム(edkdic)で用いることができるように形態素情報を抽出ためのプログラムです。

Usage:

```
mo_extr                                          ・・・・対話モード
mo_extr  分野名  形態素解析結果ファイル          ・・・・バッチモード
```

引数の説明

分野名　　$JMADATA/edoutの下にある日本語形態素解析結果ファイルのあるディレクトリを指定する。同名のディレクトリが$JMADATA/freqの下にもなければなりません。

形態素解析結果ファイル

形態素解析結果のファイルを指定します。
ファイル名はED_で始まっていなければなりません。
※　ＡＮＡの出力ファイルを入力として使う時は、名前をED_で始まるように rename することに加えて、エディタで"NOANA"を含む括弧を（中身を含めて）すべて消去しておく必要があります。

（例）

```
mo_extr  msc  ED_tel31-06
```

入力ファイルとしては$JMADATA/edout/msc/ED_tel31-06 が指定され、出力は$JMADATA/freq/mscの下にED_がFD_,FL_になったファイルが得られます。このうちFD_の方がedkdicで使われるファイルであり、FL_はFD_の内容をチェックするためのものです。

環境

環境変数JMAHOME が指定されていて、その下にはディレクトリ exec,pub,lib が存在し、それぞれの下に必要なファイル（extract.susおよび hinshi等）が存在しなければなりません。さらに、JMADATAが指定されていて、その下にディレクトリedoutとfreqが存在している必要があります。

実　行　内　容	画　　面　　出　　力	入　力　方　法
1　実行命令	T CL> ■	(run) … 2 へ
2　抽出したい ディレクトリ 選択	形態素抽出プログラム (0) jpn　　(1) key (2) new　　(3) ppr (4) tel	形態素の ディレクトリ選択 … 3 へ 空打ち … 終了
3　抽出したい ファイル選択	形態素抽出プログラム (0) ED_new-001-01 (1) ED_new-001-02 (2) ED_new-001-03 (3) ED_new-001-04 (4) ED_new-001-05 : :	ファイル選択 … 4 へ ^n … 下にスクロール ＜ ^p … 上にスクロール
4　読み込み	分野名 _ new ファイル名 _ ED_new-001-02 しばらくお待ち下さい	（操作不要）… 2 へ

mo_extrの出力ファイル例（FL_）

もしもし	もしもし	もしもし	感動詞	ー ー	ー ー
。			記　号	ー ー	ー ー

文番号 3

事務局：			記　号	ー ー	ー ー

文番号 4

はい	はい	はい	感動詞	ー ー	ー ー
。			記　号	ー ー	ー ー

文番号 5

会議	かいぎ	会議	普名詞	ー ー	ー ー
事務	じむ	事務	普名詞	ー ー	ー ー
局	きょく	局	普名詞	ー ー	ー ー
で	で	だ	助動詞	ー ー	連用
ございます	ございます	ございます	補動詞	特殊	終止
。			記　号	ー ー	ー ー

文番号 6

質問者：			記　号	ー ー	ー ー

文番号 7

会議	かいぎ	会議	普名詞	ー ー	ー ー
の	の	の	格助詞	ー ー	ー ー
宿泊	しゅくはく	宿泊	普名詞	ー ー	ー ー
施設	しせつ	施設	普名詞	ー ー	ー ー
について	について	について	格助詞	ー ー	ー ー
お	お	御	接頭語	ー ー	ー ー
尋ね	たずね	尋ねる	本動詞	下一	連用
し	し	する	補動詞	サ変	連用
たい	たい	たい	助動詞	ー ー	連体
の	の	の	準助詞	ー ー	ー ー
です	です	です	助動詞	ー ー	終止
が	が	が	接助詞	ー ー	ー ー
。			記　号	ー ー	ー ー

mo_extrの出力ファイル例（FD_）

```
(0
 #S(OUTPUT :O_DATA (161163) :O_REGULAR_HIRAGANA NIL :O_REGULAR_KANJI NIL
           :O_HINSHI 0 :O_KATSUYOU NIL :O_KATSUYOU_TYPE NIL :O_BEFORE NIL
           :O_AFTER NIL :O_PRIOR NIL :O_DICT_ID NIL :O_FREQENCY 19
           :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (188193 204228 188212 161167) :O_REGULAR_HIRAGANA NIL
           :O_REGULAR_KANJI NIL :O_HINSHI 0 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 6 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (187246 204179 182201 161167) :O_REGULAR_HIRAGANA NIL
           :O_REGULAR_KANJI NIL :O_HINSHI 0 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 4 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (161162) :O_REGULAR_HIRAGANA NIL :O_REGULAR_KANJI NIL
           :O_HINSHI 0 :O_KATSUYOU NIL :O_KATSUYOU_TYPE NIL :O_BEFORE NIL
           :O_AFTER NIL :O_PRIOR NIL :O_DICT_ID NIL :O_FREQENCY 2 <
           :O_COMMENT NIL))
(4
 #S(OUTPUT :O_DATA (197208 207191)
           :O_REGULAR_HIRAGANA (164200 164166 164237 164175)
           :O_REGULAR_KANJI (197208 207191) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 3 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (205209 187230) :O_REGULAR_HIRAGANA (164232 164166 164183)
           :O_REGULAR_KANJI (205209 187230) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 3 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (178241 181196) :O_REGULAR_HIRAGANA (164171 164164 164174)
           :O_REGULAR_KANJI (178241 181196) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 2 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (189187 189234)
           :O_REGULAR_HIRAGANA (164184 164229 164166 164183 164231)
           :O_REGULAR_KANJI (189187 189234) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 2 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (204190 193176) :O_REGULAR_HIRAGANA (164202 164222 164168)
           :O_REGULAR_KANJI (204190 193176) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 2 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (187246 204179) :O_REGULAR_HIRAGANA (164184 164224)
           :O_REGULAR_KANJI (187246 204179) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 1 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (182201) :O_REGULAR_HIRAGANA (164173 164231 164175)
           :O_REGULAR_KANJI (182201) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 1 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (187234 181222)
           :O_REGULAR_HIRAGANA (164183 164173 164229 164166)
           :O_REGULAR_KANJI (187234 181222) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 1 :O_COMMENT NIL)
 #S(OUTPUT :O_DATA (197192) :O_REGULAR_HIRAGANA (164198 164243)
           :O_REGULAR_KANJI (197192) :O_HINSHI 4 :O_KATSUYOU NIL
           :O_KATSUYOU_TYPE NIL :O_BEFORE NIL :O_AFTER NIL :O_PRIOR NIL
           :O_DICT_ID NIL :O_FREQENCY 1 :O_COMMENT NIL))
```

辞書構築プログラム

edkdicで作られた辞書（ディレクトリ）から日本語形態素解析プログラム（ＡＮＡ）で用いる辞書を構築するためのプログラムです。

Usage:

```
sdicgen  [-f]  辞書名  分野名
sdicgen   -c   辞書名
```

引数の説明

辞書名	$JMADICT/SDICの下に作られるべき辞書の名を指定します
分野名	$JMADICTの下にあるedkdicで作ったファイルのあるディレクトリを指定します
オプション	
f	中間ファイル（テキストファイル）作成の作業までを行います
c	中間ファイルから辞書を作成します

（例）

```
sdicgen  SDWK  work
```

入力としては $JMADICT/work （ディレクトリ）が指定され、出力の辞書名はSDWKでは$JMADICT/SDICの下に様々なファイル（SDWK.dic,SDWK.ix1等）が作られます。拡張子を持たないSDWKが中間ファイルでこれをエディタなどで直してから c オプションで希望する辞書を作成することもできます。
※既存の辞書ファイルと統合したいときはこの中間ファイルを統合し、UNIXのsortやuniqコマンドで整形してから sdicgen -c を実行します。

環境

環境変数JMAHOME が指定されていて、環境変数JMADICT が $JMAHOME/dict に指定されている必要があります。分野名で指定するディレクトリの下にはディレクトリHIRA,KATA,KANJ,ETCとファイルIdiom,maxfreqが存在している必要があります。また、ディレクトリSJMAHOMEの下にはディレクトリtool/src/sdic,pubが存在し、それぞれの下に必要なファイル（sdicinit.lsp,sdicgen.fasおよび hinshi等）が存在しなければなりません。

形態素関連プログラム関連図