

The logo consists of two overlapping rectangular frames. The inner frame is slightly offset to the right and down from the outer frame. The text "WinMorph" is centered within the inner frame.

WinMorph

Tutorial

Tokyo University of Foreign Studies Sano Laboratory

**WinMorph
Tutorial**

**The information in this manual is subject to change without notice and
should not be construed as a commitment by
Tokyo University of Foreign Studies.**

**Copyright (C) 1998, 1999 by Tokyo University of Foreign Studies. All
Rights Reserved.**

TERMS AND CONDITIONS FOR USE OF "ICOT FREE SOFTWARE"

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | |

The following Terms and Conditions for use of "ICOT FREE SOFTWARE"

is applied to all of:

- * ICOT Free Software that ICOT releases
- * Revised ICOT Free Software after its release
- * Software produced by AITEC's Research Funding Program.

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | |

1. Purposes and Background of ICOT Free Software.

The Institute for New Generation Computer Technology ("ICOT") had been promoting the Fifth Generation Computer Systems project under the commitment of the Ministry of International Trade and Industry of Japan (the "MITI"). Since April 1993, ICOT has been promoting the Follow-on project to the FGCS project. This follow-on project aims to disseminate and further develop FGCS technology. The FGCS project and the Follow-on project (collectively, the "Project") have been aimed at creating basic technology for novel computers that realizes parallel inference processing as their core mechanism, and contributing toward the progress of computer science by sharing innovative knowledge and technology with the research community worldwide.

Innovative hardware and software parallel inference technology has been under development through the Project, which involves varieties of advanced software for experiments and evaluation. This software, being at a basic stage of research and development, should be disseminated widely to the research community.

According to the aims of the Project, ICOT has made this software, the copyright of which does not belong to the government but to ICOT itself, available to the public in order to contribute to the world, and, moreover, has removed all restrictions on its usage that may have impeded further research and development in order that large numbers of researchers can use it freely to begin a new era of computer science.

This program together with any attached documentation (collectively, the "Program") is being distributed by ICOT free of charge as ICOT Free Software.

2. Free Use, Modification, Copying and Distribution

Persons wanting to use the Program ("Users") may freely do so and may also freely modify and copy the Program. The term "modify," as used here, includes, but is not limited to, any act to improve or expand the Program for the purposes of enhancing and/or improving its function, performance and/or quality as well as to add one or more programs or documents developed by Users of the Program.

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.

3. 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.

We are the members of the project.

SANO, Hiroshi

TAKAHASHI, Sakutaro

Xie Yun

Kim Mi Song

INOUE, Yoshikatsu

MAEJIMA, Megumi

FUJIMOTO, Yuka

WinMorph / Tutorial

February 1999

Tokyo University of Foreign Studies Sano Laboratory

Printed in Japan.

1 On this document

This tutorial features Winmorph :

Winnmorph is a workbench to create rule files for morphological analysis, which uses a morphological analysis engine (Breakfast(R), Anima(R)).

In this tutorial, steps to make up morphological analysis application which is executed by Winmorph, and how to create morphological analysis rules and dictionaries are provided as the subject.

1.1 Prerequisites

This tutorial is designed for those who have never used Winmorph before. Comprehension of the conception of morphological analysis and knowledge of these analysis rules or dictionaries are not required.

If you are new to the use of GUI based applications, refer to your platform (especially your operating system) manuals, in particular to know about dialogical operations of the application with a mouse .

1.2 Documentation Roadmap

- *Chapter 1: On This Document

- *Chapter 2: Installing Winmorph

- *Chapter 3: The Basic Operations of Winmorph
(Starting, Quitting and Window)

 - The User Interface of Winmorph etc.

- *Chapter 4: Morphological Analysing with Winmorph

- *Chapter 5: Explanation of the Conception of Morphological Analysis
(of Japanese text) and the way of making up morphological
analysis rules and building dictionaries

- Appendix (Glossary):

 - Technical Terms used in this tutorial and the interface of
Winnmorph

1.3 Conventions Used in the Document

This chapter explains conventional rules used in the document to distinguish technical terms, the structure of computer language, mouse buttons, and operations by keyboard.

Font Styles

Several font styles are used for emphasis throughout the document for different purposes. They are as follows :

Examples	Description
.....	

Special Symbols

They are as follows :

Examples	Description
.....	
注意 (Caution)	describes information ; to ignore this might cause the loss of data.
警告 (Warning)	describes information ; to ignore this might cause damage to the system.

Rules for the Screen

Hard copies of the screen will be used to tell the results of various operations as the samples.

Mouse

Terms to describe operations with a mouse are as follows :

Notations in this document	Description of operations
click	Press the left button of the mouse and release.
double-click	Press the left button of the mouse twice, not moving the pointer.

1.4 Other resources

You can also use the following references and manuals as well :

Online documents

PDF documents can be referred using Adobe Acrobat Reader. Web documents can be referred to using your favourite web browser.

2 Installing WinMorph

This chapter covers the following information :

- è Installing Winmorph
- è Installing Breakfast
- è Extracting wjpmph15
- è Setting up VisualBasic5.0
- è Setting up VisualBasic5.0 ServicePack3

2.1 Setting Up the Environment

Information in this section is provided in connection with platform system for using Winmorph.

(1) Downloading Winmorph(in Win Zip format)

Winmorph is distributed in ZIP format archive, which can be extracted with WinZip(R) (shareware from Nico Mak Computing, Inc.). You can download Win Zip from :

<http://www.winzip.com/>

The site to load Winmorph for free is :

<ftp://ftp.icot.or.jp/>

(2) (a) Extracting with Win Zip

- 1 Start File Manager or Explorer. Double-click \Winmorph.zip" to start Win Zip.
- 2 Click the \Open" button to open the \Open Archive" window.
Select Winmorph.zip and open the file.
- 3 Click the \Extract" button to open up the \Extract" window.
- 4 Specify the folder you want the archive extracted into. For example, answering \C:\morph" will extract the files into the \morph" folder on your C drive.
Click the [Extract] button. If Winmorph is successfully extracted, the C:\morph folder will contain 23 files, as \AsycFilt.dil", \GAPI32.dil", \winmorph.exe", etc.
- 5 Close Win Zip.

(b) Setting Up Winmorph

- 1 Double-click "Setup.exe" included in the Winmorph folder created in 4.1. Click [OK] to open the set-up window.
- 2 Click the upper right button in the [Win Morph セットアップ (set-up Winmorph)] window to start the Set-up wizard.
- 3 Set-up is completed.

(3) Installing Breakfast You can download Breakfast from :

URL <http://www.fujitsu.co.jp/hypertext/breakfast/download.html>

Here you can also check information on how to install the software.

(4) (a) Extracting wjpmph15

wjpmph15 is the IFS general Japanese morphological analysis grammar rule, which was developed at Tokyo University of Foreign Studies. LAX(morphological analysis system), part of the ICOT free software, was ported and revised for JUMAN 2.0 (morphological analysis system released from Matsumoto Laboratory, Nara Institute of Science and Technology). On this basis, wjpmph15 was built up, which also includes new data as a dictionary of analysis.

This morphological analysis grammar rule is able to analyse Japanese subsets morphologically. The system of the structure of form, which this morphological analysis rules requires on the condition, does not depend on the system of parts in structural grammar as much as possible. Thus the analysis result is highly independent, and can be widely used from language research to application development.

You can download wjpmph15.tgz from :

URL <http://www.fujitsu.co.jp/hypertext/breakfast/download.html>

The file is archived with TGZ (tar g-zipped) format, which can be extracted with Win Zip. Successfully extracted, folders and files, e.g. "dic", "dict", "rule", "README.txt" will be created.

(b) Copy the "dic" dictionary

Copy the "dic" folder extracted in 4.1 to the Winmorph folder created in 2.1.

Using Explorer, you can do this by the "drag and drop" operation.

Point to the "dic" folder. Press the right button of the mouse, and before releasing it, move the pointer to the Winmorph directory. Releasing the button will open up the menus. Choose the "Copy" key.

To Modify the Source Code

Information in this section includes the maintenance of your platform system to modify the source code for Winmorph.

Winnormorph is developed with Microsoft(R) Visual Basic(R) 5.0. In order to modify the code, you need to set-up Visual Basic 5.0.

Follow the steps as described below to maintain the developmental system :

è Setting up VisualBasic5.0

To use the Winnormorph application, set-up \VB5.0". Select the \標準 (standard) set-up" in this installation.

è Apply Visual Basic Service Pack 3

You can download Visual Basic Service Pack 3 from Microsoft's web site.

Find this service pack at the following location :

Microsoft Corporation

URL <http://www.microsoft.com/>

Clicking the self extracting file will extract files in a folder named \Vs97_sp3".

See such folders or files as 'all', 'deu', 'seup.inf', 'seup.exe' created in the \Vs97_sp3" folder.

Clicking \Setup.exe" in the \Vs97_sp3" folder will apply Service Pack 3 to Visual Basic 5.0.

Click \Help" ! \Version" on the Visual Basic menus to check whether Service Pack 3 is correctly applied or not. If the installation is succeeded, you can see 'SP3' on your display when you set-up \VB5.0".

è Winnormorph (source code)

The source code of Winnormorph is distributed in ZIP format archive, which can be extracted with WinZip(R).

3 Basic Operation and Interface

This chapter covers the following information :

è Starting Winnormorph

è The Composition of Winnormorph Window

è The Usage of the Interface for Winnormorph

è Quitting Winnormorph

3.1 Relating Winmorph...

Winmorph is a workbench to create rule and dictionary files for morphological analysis of Japanese text with using a morphological analysis engine named Breakfast(R). Winmorph enables you to modify the present reference to the rules for morphological analysis, delete or add one/some of the existing dictionaries Winmorph already has, as well as to create new references of them on your own.

With the following function of Winmorph, you can make up very easily some other individual rules or dictionaries for executing this analysis.

- è Creating rules and dictionaries for morphological analysis :
- è Modification for the current rules or dictionaries, and addition to them :
- è Compiling and creating a package of rules and dictionaries, and testing analysis by some optional package of them :

3.2 Setting up Winmorph

The instructions in this chapter assume that Winmorph and Breakfast are already installed on your platform. You should note this in advance of the set-up of Winmorph. <download: Breakfast>

<http://www.fujitsu.co.jp/hypertext/breakfast/download.html>

1. Select and double-click the 'winmorph.exe' file (ex. Path: C:\morph\winmorph.exe).
2. Or in another way, set-up from Windows platform as follows :
 - (a) Click the 'Open Selected File' menu in the task bar 'start' menu, and enter the path to 'winmorph.exe' file (ex.Path: C:\morph\winmorph.exe).
 - (b) Click the 'Enter' button, and Winmorph will be set-up and displayed.

Notes	
	Make sure the path name of the file 'winmorph.exe' is the same one as when you set for the 'extraction of Winmorph' before.

3.3 The Outview of Winmorph

Set up Winmorph, and the Winmorph window as follows (Figure 1) will come up on your display.



図 1: Winmorph window

3.4 Winmorph window

Winnorph window is shown as (figure 2):

This window has 4 regions of 'title bar', 'menu bar', 'tool bar', and 'working region for text to input and output'.

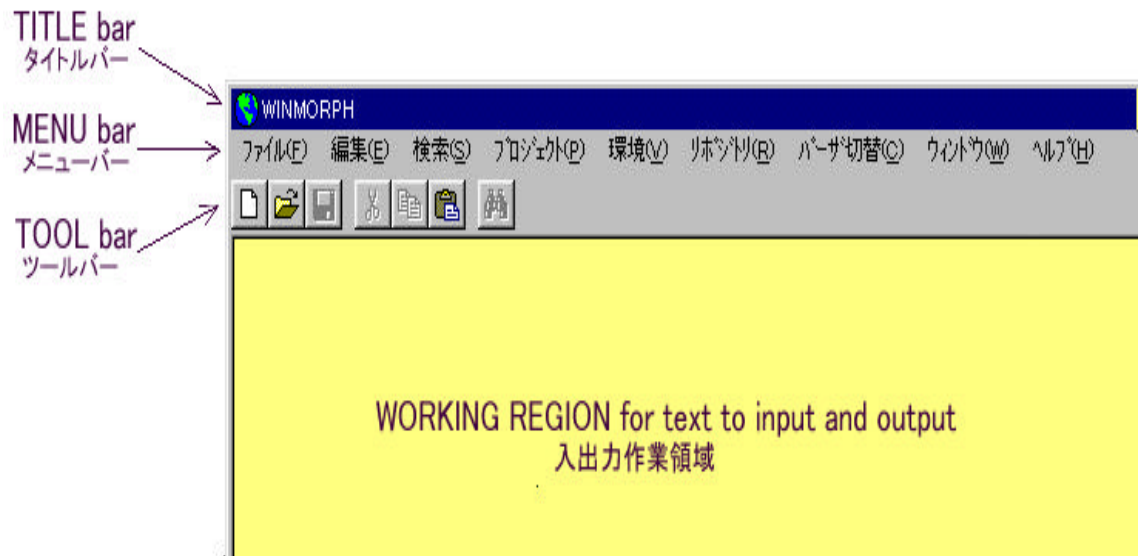


図 2: Winmorph window(2)

3.4.1 Window Menus

Selecting one of the menu functions in the menu bar, you can access each function of Winmorph.

The summary of the context of the menu functions is described in the following chart :

MENUS	SUB MENUS	COMMANDS
ファイル FILE	新規作成 NEW OPEN	Opens 2 windows; the upper is 'text window', and the lower is 'run analysis window'.
	開く OPEN	Shows the [ファイルを開く]\Open File" dialogue box, and then opens the selected file. To open a file, you can also 'Drag & Drop' your favourite text or file onto the upper text window with file manager or explorer.
	上書き保存 SAVE	Saves an editorial file as the same name.
	名前をつけて保存 SAVE AS	Shows the [名前を付けて保存]\Save As" dialogue box, and saves an editorial file as the file name entered in the box.
	印刷 PRINT	Shows the [印刷] \Print" dialogue box, and prints the selected range of the text in an editorial file.
	エディタを閉じる CLOSE(CLOSE WINMORPH)	Closes the application of Winmorph, with a message inquiring to/not to save the text of an editorial file.
編集 EDIT	元に戻す UNDO	Undoes (Cancels the previous 1 edit operation).
	切り取り CUT	Cuts the selected text, and it copies on clipboard.
	コピー COPY	Copies the selected text on clipboard.
	貼り付け PASTE	Pasts the text copied on clipboard at the point of the cursor.
	削除 DELETE	Deletes the selected text.
	すべて選択 SELECT ALL	Selects all the range of the temporary text in edit.
検索 SEARCH	検索 SEARCH	Shows the [文字列の検索] \Search String" dialogue box, and finds out the string entered in the box.
	次を検索 SEARCH THE NEXT	(almost the same function as above) Finds the next string.

MENUS	SUB MENUS	COMMANDS
プロジェクト PROJECT	プロジェクト定義 DEFINE PROJECT	Shows the [プロジェクト定義] "Define Project" dialogue box; de- termine some 'rule' or 'dictionary' files.
	パーザ作成 CREATE PARSER	Shows the [パーザの作成] "Create Parser" dialogue box, and creates a parser based on the deter- mined 'rule' and 'dictionary' files.
	パーザ実行 EXECUTE PARSER	Analyses the text in text window with the created parser.
環境 SYSTEM	環境変数設定 SET VALUABLE SYSTEM	Sets system : the system on setting the path of Breakfast, how to display the analysis, etc.
	ツールバー表示 SHOW/HIDE TOOL BAR	Check here, and shows or hides the tool bar.
	ステータス表示 SHOW STATUS	Shows the temporary editorial status at the bottom of the Win- morph window.
リポジトリ REPOSITORY	規則ファイル登録 REGISTER RULE FILE	Shows the [規則ファイルの選択] "Select Rule File" dialogue box ; you can indicate rule files.
	辞書ファイル登録 REGISTER DICTIONARY FILE	Shows the [辞書ファイルインデッ クスの登録]"Register Dictionary In- dex File" dialogue box ; you can in- dicate dictionary files.
	リソースファイル登録 REGISTER RESOURCE FILE	Determine a resource file.
パーザ切替 SWITCH PARSER		Switches 'chasen' or 'breakfast' to the other parser.
ウィンドウ WINDOW		You can select a window between text window and run analysis win- dow for edit.
ヘルプ HELP		Opens the guidance and online documents.

3.4.2 TOOL BAR

You may access easily by tool bar button, besides by using the menus in the menu bar of window, to each frequently run function for Winmorph.

The summary of the buttons is as below :

3.4.3 Windows to Execute Analysis of Text and Display Analysis

Working Region for Text to Input and Output The basic function of morphological analysis is targeted at running analysis and showing the result of the analysis.

Working Region for Text to Input and Output is the region of two windows to input text and output the result of that analysis.

By 'drag & paste', you can paste text from other application to the Winmorph window for inputting text ; as well, it is possible to paste text from the window for outputting, as the result of the run analysis, to other application.

3.5 Quitting Winmorph








To quit Winmorph, click the [ファイル]\File" menu [終了]\Close" button in the window.

4 Analysing with Winmorph

This chapter covers the following information :

execution of morphological analysis with Winmorph

- è To use the morphological analysis functions of Breakfast interactionally through Winmorph
 - è To make up software for morphological analysis with the 'wjpmph15' dictionary and execute morphological analysis
 - ° You need a morphological engine called 'Breakfast' and a morphological dictionary, 'wjpmph15', which are already installed to start execution of morphological analysis with Winmorph.
- Both are available for download from the web page below :

BUTTONS	COMMANDS	FUNCTIONS
	新規作成 NEW OPEN	Opens 2 windows for text to input and output for the analysis.
	開く OPEN	Opens an exiting file.
	名前をつけて保存 SAVE AS	Saves a file as the same name.
	切り取り CUT	Cuts the selected text.
	コピー COPY	Copys the selected text.
	貼付け PASTE	Pastes the cut or copied text.
	検索 SEARCH	Finds out the string entered in the box.

<http://www.fujitsu.co.jp/>

(in this site, the instructions for installation are also covered)

4.1 The Instructions of the Morphological Analysis

There are two ways to run morphological analysis :

- (a) Using the existing morphological analysis parser compiled before, to execute analysis
- (b) Using the morphological analysis parser created by yourself with some morphological rules and dictionaries to execute analysis

4.1.1 Preparation

Click "\Set-up Valuable System(環境変数設定)" in the menu bar, and enter the path to the "\cbf.exe" file (executive binary file of Breakfast).

For instance, type "\C:\Program Files\Fujitsu\404\bin\cbf.exe" as the full name of the file.

4.1.2 Using the previously compiled parser in morphological analysis

To run analysis through Winmorph using the compiled parser, follow the instructions as described below :

- 1 Click "\winmorph.exe" in file manager or explorer, and set up Winmorph.
- 2 Select [新規作成 Open new(N)] in the menu bar [ファイル File(F)] function and open two working windows in the working region for input and output of text.

Text window for input is located above, and run analysis window for output is below.

- 3 Enter Japanese text to analyse in text window. Or drag and drop text from other application and paste it there, if desire.
- 4 Select [パーザの実行 Execute parser(P)] in the menu bar [プロジェクト Project(P)] function.

The resulting analysis is output in run analysis window after executing the analysis.

- 5 To save the data of the text :

Click the one window consisting of your favourite text to save.

For example, if you want to save the result of the analysis, click run analysis window showing the analysed text.

Select [上書き Save(S)] or [名前をつけて保存 Save as(A)] in the menu bar [ファイル File(F)] function.

Without clicking directly on the window, you may click and chose the desirable window for edit by checking the name listed in the menu bar [ウィンドウ Window(W)] function with your mouse.

Instead of typing text in text window, if you use an existing text file, already edited and saved, see the steps as follows :

1. Select [新規作成 Open(O)] in the menu bar [ファイルFile(F)] function (as the step [2] shown before).
2. Sort a file in the dialogue box and press the Open(開く) key.

And text window of which title bar is by the path name of the selected file will be displayed in the upper side of the Winmorph window.

4.1.3 Compiling your own parser which consists of morphological rules and dictionaries

This section features how to compile a parser with morphological rules and dictionaries and execute analysis by Winmorph with the parser for receiving the result of the analysis.

To run analysis through Winmorph using the individual compiled parser, follow the steps as described below :

1. Click "winmorph.exe" , and set up Winmorph.
2. Register rule files by clicking [規則ファイル登録 Register rule file(R)] in the menu bar [リポジトリ Repository(R)] function and sort the files.
3. In this sort, type the path to "rc file", a classification dictionary file (as to parts of speech), and other files.

Or indicate the file paths by clicking the [参照 Reference] button and selecting each corresponding rule file.

For instance, in case of "wjpmph15", if you make reference to and select the "humanrc" file from the "dic" folder as a "rc file" and click the [開く Open] key in the dialogue box, all the corresponding files will be set automatically with each path.

You will confirm all the paths displayed in the boxes in the reference dialogue window.

} To resist files, each user can select each file, and click the [登録 Register] button in the dialogue.

4. According to the message prompt, type the name of the registered files, and click the [登録 Register] button in the message box.

A new file named \"(file name).rif\" will be created in the ¥dic folder.

5. Select [辞書ファイル登録 Register dictionary file(D)] from the menu bar [リポジトリ Repository(R)] menu, and register dictionary files.

Indicate the paths of your favourite dictionary files as reference in the same way as when you chose rule files before.

By clicking [辞書ディレクトリ変更 Change directory for dictionary], you can refer to files and open the dictionary files selected there to set the registry.

In this process, when you indicate \"¥dic\" and have all the included dictionary files shown in the left side of the window, you should select favourite files by clicking on them and just press the [追加 Add] or [全追加 Add all] key ;

You will confirm those selected files displayed in the right side of the window.

} To correct operations, press the [削除 Delete] or [全削除 Delete all] key.

6. According to the message prompt, type the name of the registered files, and click the [登録 Register] button in the message box.

A new file named \"(file name).dif\" is created in the ¥dic folder.

7. Select [プロジェクト定義 Define project(P)] from the menu bar [プロジェクト Project(P)] menu and type the path to the [規則ファイルディレクトリ / 規則インデックスファイル Rule file directory / rule index file].

(The [登録 Reference] button will help you select the index file created before ; normally, you should indicate the ¥dic folder including such a file as \"jumanrc'.)

Operate with the [追加 Add] and [削除 Delete] key in selection of [辞書ファイルディレクトリ / 辞書インデックスファイル Dictionary file/dictionary index file] (This equals the selection of the already created index files).

After finish of all these definition, press the [OK] key.

8. Click [パーザ作成 Create parser(C)] in the menu bar [プロジェクト Project(P)] and press the [実行 run] key to compile the registered rule and dictionary files.

Notes	
	It takes a long time until the completion of creating a parser.

9. After finish of the compilation of the selected rule and dictionary files, kind of new software for morphological analysis is created.

By now, follow the instructions as described in the former section to use the already compiled parser, and analyse desirable text morphologically through Winmorph.

5 Conception of Morphological Analysis

This chapter covers the following information:

the conception of morphological analysis and the function of software for this analysis

- è Morphological Analysis and Computer Software

- { On Morphological Analysis

- É Morpheme

- É goki, haseiji, kussetsuji

- { On Software for Morphological Analysis

- è Explanation of the Working Concept

- è Explanation of the Functional Concept

- è Explanation of the Part of the Morphological Analysis-Driving Data and Creating the Data

5.1 Morphological Analysis and Computer Software

On Morphological Analysis

Morpheme Dividing words will lead to parts called "morpheme".

Morpheme means the smallest divided linguistic form.

Linguistic explanation of morpheme follows below, quoting basically "日本文法体系論 (Systematic Treatise of Japanese Grammar)" by Kenzo Morioka.

Morphemes can be divided into two sorts in the light of the form. (1) One is called 'free form' which is pronounced solely as the division from other forms. (2) The other is 'bound form' which is pronounced serially combining with others.

For example : in the case of a sentence as "単語を分割してゆくと (tango wo bunkat-sushite ikuto : dividing words ...)"

単語を (tango wo), 単語 (tango), 分割してゆくと (bunkatsushiteikuto), 分割して (bunkat-sushite), etc. are classified in free form. And を (wo), してゆく (shiteiku), ゆくと (ikuto), と (to), etc. are bound form.

語基 goki、派生辞 haseiji、屈折辞 kussetsuji ! goki { basis(radical,theme), haseiji { derivative ending(derivation), kussetsuji { inflexion(inflexional suéx)

° goki `goki" is a morpheme which is a basic word form to derive or inflect others : like `単語 tango' and `分割す(る) bunkatsusuru' as described above. cf.`goki" is a smaller basic unit than the stem of a word.

° haseiji `haseiji" is a morpheme to combine with `goki" and derive other words. `(し)てゆ(く)(と) (shi)tei(ku)(to)' as above ; preáx, ináx, or suéx corresponds to this.

° kussetsuji `kussetsuji" is a morpheme to combine with goki and create a `kussetsugo"(inflected word). Postpositional particle and inflected conjugation or comparison like `(単語)を (tango)wo', `(分割し)て (bunkatsushi)te', and `(ゆ)く (i)ku' as shown before are equal to this.

For example : a sentence of `単語を分割してゆくとそれ以上分割できない最小の形式に達する。

(tango wo bunkatsusuruto koreijyou bunkatsudekinai saishouno tann-i ni tassuru. - Dividing words will lead to parts which cannot be divided anymore.)" `単語 tango', `を wo', `分割す(る) bunkatsusu(ru)', `ゆ(く) i(ku)', `と to', `それ sore', `以上 ijou', `分割でき(る) bunkatsudeki(ru)', `な(い) na(i)', `最小 saishou', `の no', `形式 keishiki', `に ni', `達す(る) tassu(ru)', and `.' are all morphemes.

These morphemes are listed in the table below :

morpheme	categories
-----+-----	
「単語 tango」	goki
「を wo」	kussetsuji
「分割す(る) bunkatsusu(ru)」	goki
「て te」	haseiji
「ゆ(く) i(ku)」	haseiji
「く ku」	kussetsuji

「と to」	kussetsuj i
「それ sore」	goki
「以上 ijou」	goki
「分割でき(る) bunkatsudeki (ru)」	goki
「き ki」	hasei j i
「な(い) na(i)」	goki
「い i」	kussetsuj i
「最小 saishou」	goki
「の no」	kussetsuj i
「形式 keishiki」	goki
「に ni」	kussetsuj i
「達す(る) tassu(ru)」	goki
「る ru」	kussetsuj i
「。」	goki

-----+-----

It is characteristic in Japanese that each morpheme combines with others in a series and text has no definite end of a word or partition of a sentence .

On Softwares for Morphological Analysis

Software for morphological analysis is one to analyse text morphologically by computer.

It separates a string entered into a computer, and divides more into a series of morphemes, referring to each index of morphological dictionaries. This description where

the appropriate ways of division into morphemes are written is called 'morphological analysis rules'.

In such a word as \を wo" for instance, \単語 tango" connects with the word \を wo", but \分割す(る) bunkatsusu(ru)" is not to be followed by it.

Therefore, morphological analysis consists of those parts :

1. to divide words or control the way of division
2. to judge as to whether the divided strings are registered in morphological dictionaries

3. to consult as to whether the connections of the divided morphemes are proper
We call the thing of 1 'a morphological analysis engine', and a set of 2 and 3 'morphological analysis - driving data'.

5.2 Explanation of the Working Concept

Morphological analysis software works as when text strings are input, it outputs the strings of the morphemes of the text as the resulting analysis.

The working concept of the software is as described in the following figure :

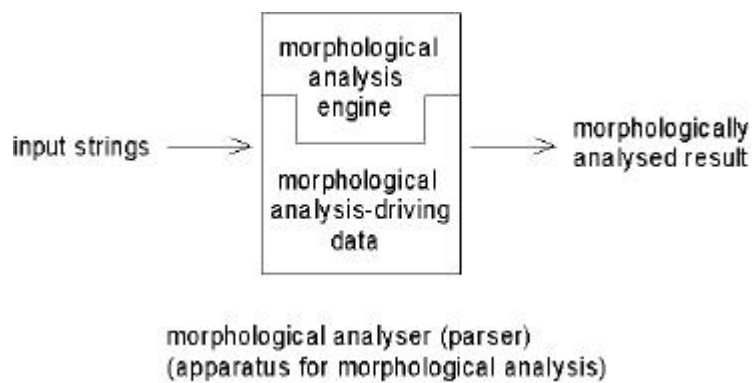


図 3: Software for Morphological Analysis

A 'parser'(morphological analyser) is the one that puts together a morphological analysis engine and morphological analysis-driving data.

5.3 Explanation of the Functional Concept

The basic function of software for morphological analysis is to change input strings into the strings of the morphemes.

On the assumption that it analyses the following input text :

Input string : 「日本語の形態素解析の機能を説明します」
\nihongo no keitaisokaiseki no kinou wo setsumeishimasu"

The resulting analysis is as follows :

Executed result : -日本語-の-形態素-解析-の-機能-を-説明-し-ま-す-

\-nihongo-no-keitaiso-kaiseki-no-kinou-wo-setsumei-si-ma-su"

It provides morphological dictionary data concerning the classified names of morphemes, how to read(sound) them, etc., as appendix.

morpheme	classified name	reading
-----+-----+-----		
「	開き括弧 hirakikakko	ひらきかっこ hirakikakko
日本語	体言語基 taigengoki	にほんご hi hongo
の	連体助辞 rentaijoi	の no
形態素	体言語基 taigengoki	けいたいそ keitaiso
解析	体言語基 taigengoki	かいせき kaiseki
の	連体助辞 rentaijoi	の no
機能	体言語基 taigengoki	きのう kinou
を	格助辞 kakujoji	を wo
説明	体言語基 taigengoki	せつめい setsumei
し	用言語基「する」 し yougengoki`suru'	
ま	丁寧接辞「ます」 ま teineisetsuji`masu'	
す	活用助辞「す」 す katsuyoujoji`sus'	
」	閉じ括弧 tojikakko	とじかっこ tojikakko

-----+-----+-----

3 cf.

`hirakikakko',`tojikakko' : quotation mark, bracketing

`taigengoki' : goki for nouns(indclinable words that can stand as kind of the subject of a sentence),(nominal goki)

`yougengoki' : goki for verbs, adjectives,(declinable words), (verbal goki)

`rentaijoji' : aÉx for connection with nouns

`kakujoji' : aÉx for cases

`teineisetsuji' : aÉx for expression of politeness

`katsuyoujoji' : aÉx for conjugations

`fukuyougengoki' : goki for adverbials

5.3.1 Part of Morphological Analysis-Driving Data

The part of morphological analysis-driving data memorizes the information (morphological dictionary data and morpheme connection rules) for a morphological analysis engine to transform text strings into the connected morpheme strings by using that.

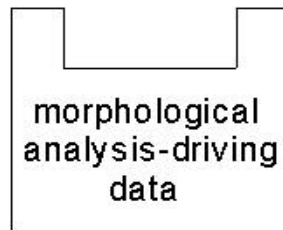


図 4:

In order to make software for morphological analysis, generally, we do this by ourselves; we create data files in a form as it includes comprehensive knowledge of mother tongue(Japanese) Studies and linguistics, and likeliness of people's understanding.

That is, we need to write files of which description is on data of morphological dictionaries and morpheme juncture rules for analysis.

Moreover, those are changed by transformational software into the data form to be used in a morphological analysis engine. With a set of morphological analysis-driving data created in this way and a morphological analysis engine, we are enabled to execute morphological analysis.

The way a parser is composed, the main functional part of morphological analysis software, can be explained by means of a diagram, as below :

[1] Make a file using text editor.

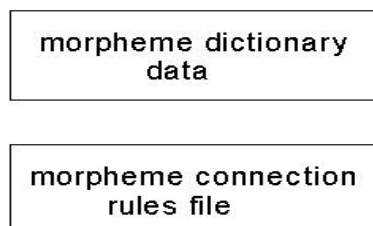


図 5:

[2] Convert the data format of the file.

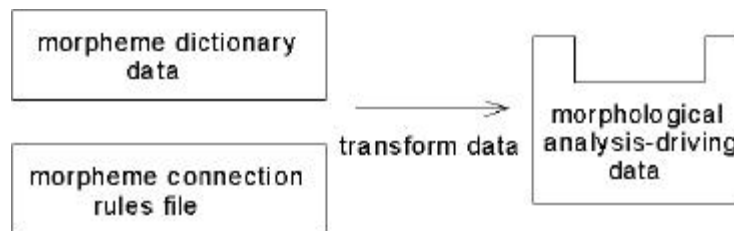


Fig 6:

[3] Set a parser going (Create a parser and operate).

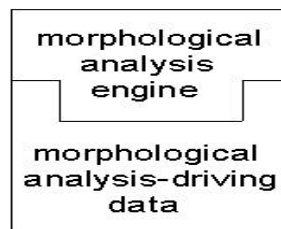


Fig 7:

[4] Input text for morphological analysis

5.4 Creating the Data

5.4.1 Morphological dictionary data

On the method of creating morphological dictionary data, we write down indices in particular to a morpheme and repeat this for each morpheme.

Basic indices of morphological dictionary data are the entry form of the morpheme(head form), the classified name of the morpheme, the reading(sound), and other information.

The following description is structurally about those indices :

ì	_____ Basic Index of Dictionary Data _____	è
	(a) Description of Morpheme(Entry of Dictionary)	
	(b) Classified Name of Morpheme 1,(Classified Name2,3..)	
	(c) Reading	
	(d) Other Dictionary Information	
í	_____	é

For example :

\単語 tango"

- (a) Description of Morpheme
単語
- (b) Classified Name of Morpheme
体言語基 taigengoki
- (c) Reading
たんご tango
- (d) Other Dictionary Information
{

\分ける wakeru"

- (a) Description of Morpheme
分け
- (b) Classified Name of Morpheme
用言語基 yougengoki, ウ系弱変化 ukei jakuhenka
- (c) Reading
わけ wake
- (d) Other Dictionary Information
{

The precise data format differs from each morphological analysis software.

5.4.2 Creating morpheme juncture rules

Morpheme juncture rules are data which describe how the appropriate connections of morphemes are (in other words, it means which morpheme follows which morphemes).

Taking a morpheme as an example, the possibility of its connection as followed by other morphemes is all covered.

We can give the conceptual figure of the junctures as described below for instance :

morpheme A -- juncture -- morpheme B

morpheme A -- juncture -- morpheme C

...

In another way, it can be written as connection matrix like this :

	A	B	C	D ...
morpheme A	x	x	x	x
morpheme B	o	x	o	x
morpheme C	o	o	...	
morpheme D	x	x	...	
...				

The data format differs from each software in details.