



WinMorph

チュートリアル

東京外国語大学 佐野研究室 編集

WinMorph

チュートリアル

本書の内容は予告なく変更することがあります .

Copyright (C) 1998, 1999 by 東京外国語大学 外国語学部. All Rights Reserved.

(無断転載を禁ず)

[利用条件]

ICOTフリーソフトウェア（IFS）の利用にあたって

|||||||

以下の利用にあたっての条件は、

* ICOTの公開したICOTフリーソフトウェア* 公開後改良されたICOTフリーソフトウェア* AITECの委託研究によって作成されたソフトウェアすべてに対して適用されるものである。

|||||||

ICOT 無償公開ソフトウェアの利用条件

1. ICOT 無償公開ソフトウェアの目的

財団法人新世代コンピュータ技術開発機構（以下、ICOTという）は、日本国通商産業省より委託され、第五世代コンピュータ・プロジェクトを推進してきた。また、平成5年度からは、このプロジェクトの後継プロジェクトとして、第五世代コンピュータの研究基盤化プロジェクトを推進している。第五世代コンピュータ・プロジェクトおよびその後継プロジェクト（以下、これらの一連のプロジェクトを本プロジェクトという）は、並列推論処理を中核メカニズムとする新しいコンピュータの基礎技術を創出し、その知見と技術を世界の研究者と共有することによって、コンピュータ科学の発展に貢献することを目的としている。

本プロジェクトによって、並列推論マシン、並列推論ソフトウェア技術といった新しい技術が開発され、また、こうした技術開発に伴い、多くの先進的なソフトウェアが試作されている。これらのソフトウェアは、基礎的な研究開発段階にあるため、多くの研究者に広め発展させていくべきものである。

そこで、ICOTは、本プロジェクトの国際貢献の目的に鑑み、著作権が国ではなくICOTに帰属することとなるこれらの研究開発段階のソフトウェアを、「ICOT 無償公開ソフトウェア」として公開してきた。これらのソフトウェアについては、研究開発のための障害となるいっさいの制約をはずすことによって、多くの研究者の方々に自由に利用してもらい、新しいコンピュータ科学への貢献を实践したいと考えている。

本プログラム及びドキュメント（以下、本プログラムという）は、「ICOT 無償公開ソフトウェア」の一つとして、ICOTにおいて無償で配布しているものである。

2. 使用、変更、複製、配布の自由

本プログラムの利用者は、その使用、変更、複製を自由に行うことができる。ここでいう変更には、本プログラムの機能、性能、品質を向上させる

ために改良、拡張を行うこと、もしくは自ら開発したプログラムやドキュメントを本プログラムに追加することが含まれるが、それだけには限定されない。

本プログラムの利用者は、本「ICOT 無償公開ソフトウェアの利用条件」第3項(「無保証」)が記されていることを条件として、関連法令に違反しない限り、本プログラムそのもの、または本プログラムの変更版を第三者へ自由に配布することができる。

3. 無保証

本プログラムは、本プロジェクトの研究開発の試作物を『あるがまま』の状態を提供するものである。このため、明示的であるか黙示的であるか、または法令の規定により生ずるものであるか否かを問わず、一切の保証をつけないで提供されるものである。ここでいう保証とは、プログラムの品質、性能、市場性、特定目的適合性、および他の第三者の権利への無侵害についての保証を含むが、それに限定されるものではない。

本プログラムの利用者は、本プログラムが無保証であることを承諾し、本プログラムが無保証であることによるすべてのリスクを利用者自身で負うものとする。

従って、利用者が本プログラムを利用したこと、または利用できないこと、もしくは本プログラムを利用して得られた結果に起因する一切の損害について、著作権者である ICOT および本プログラムの開発に関与した関連機関並びにそれらの役職員及び従業員は、そのような損害の発生する可能性について、知っていたか否かにかかわらず、何らの責任も負わない。本プログラムの利用者は、本プログラムの利用を開始したことによりこれを承諾しているものとみなされる。ここでいう利用とは、本プログラムの使用、変更、複製、配布、二次的著作物の作成を含むがこれらに限定されない。

利用者が本プログラムそのもの、または本プログラムの変更版を、ICOT 以外の第三者から配布を受けた場合においても、配布を行った第三者が独自に特別な保証を文書で行わない限り、配布を行った第三者は、その利用者に対して、本プログラムに係る限りにおいて同様に何らの責任を負わないものとする。

プロジェクトメンバー

佐野 洋

高橋作太郎

Xie Yun

金美仙

井上義勝

前嶋芽久美

藤本有香

WinMorph / チュートリアル

1999 年 2 月 吉日

東京外国語大学 佐野研究室 編集

Printed in Japan.

1 本書について

このチュートリアルは、**Winmorph** について説明します。**Winmorph** は、形態素解析エンジン (**Breakfast(R)**, **Anima(R)**) を使用した形態素解析規則作成のためのワークベンチです。

このチュートリアルは、**Winmorph** で実行することのできる形態素解析アプリケーション作成の各ステップについて説明します。この説明の過程を通じて形態素解析規則と形態素辞書の構築法について説明します。

1.1 対象読者

このチュートリアルは、**Winmorph** を初めて使用する方を対象にしています。このチュートリアルの前提として形態素解析についての概念、形態素解析規則、形態素辞書についての知識は必要ありません。

グラフィカルインタフェース (GUI) を持つアプリケーションについて使用経験がない場合、特にマウスを使用してアプリケーションを対話的に操作する方法については各プラットフォーム (とくに利用コンピュータのオペレーティングシステム (OS)) の資料を参照してください。

1.2 本書の構成

このチュートリアルは次の構成になっています。

第1章 本文

第2章 **Winmorph** のインストールの仕方を説明します。

第3章 **Winmorph** の開始と終了、**Winmorph** ウィンドウの使い方、ユーザーインタフェースの対話的な操作方法を説明します。

第4章 **Winmorph** を使った形態素解析の仕方について説明します。さらに **Winmorph** で利用している形態素解析エンジンについての説明をします。**Winmorph** が提供するインタフェースを利用しないで形態素解析エンジンを使う方法についても示します。

第5章 (日本語) 形態素解析についての概念的な説明をします。形態素解析規則の作成の仕方や、形態素解析辞書の構築の仕方についても説明します。

付録 (用語集) この付録では、このチュートリアルと **Winmorph** でインタフェースで使用する用語について説明します。

1.3 規則

この節では、専門用語、コンピュータ言語の構文、マウスボタン、マウスおよびキーボード操作を区別するために、本書で使用されている表記の規則について説明します。

文字規則

本書では特別な用語を示すために次のフォントを使います。

例	使用例
.....	

特殊記号

本書では、ある特殊の項目あるいは関係を示すために次の記号を使います。

例	説明
.....	
注意	注意には、無視するとデータの損失を起こす可能性のある情報が記述されています。
警告	警告には、無視するとシステムに損傷を与える可能性のある情報が記述されています。

画面規則

さまざまな作業の結果を表すサンプル画面を使います。

マウスの操作

マウスボタンを使って操作を示すために本書で使われている用語を示します。

本書の表記	操作説明
クリック	左ボタンを押して放します
ダブルクリック	ポインタを動かさずに左ボタンを2回押して放します

他の情報源

このチュートリアルのに次の資料やマニュアルを利用することができます。

オンラインドキュメンテーション

オンラインドキュメントをブラウザーで表示するには、PDF 文章の場合、**Adobe Acrobat Reader** を利用します。WEB ドキュメントの場合には、**WEB** ブラウザーを利用します。

2 インストール方法

第2章では、Winmorph のインストール方法について説明します。

- è Winmorph のダウンロードと展開
- è Breakfast のダウンロードと展開
- è wjpmph15 のダウンロードと展開
- è VB5.0 のセットアップの方法
- è VBSP3 のダウンロードとセットアップの方法

2.1 利用環境の整備

この節では Winmorph を利用するためのプラットフォーム環境の説明をします。

2.1.1 Winmorph のダウンロードと展開

Winmorph は ZIP 圧縮形式で提供されます。WinZip (R) (Nico Mak Computing, Inc.) を使って展開します。なお、WinZip は下記のサイトからダウンロードすることができます。

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

2.1.2 WinZip を使った展開

WinZip を使った展開の手順を説明します。

1. ファイルマネージャーもしくはエクスプローラを開き、Winmorph.zip をダブルクリックし、WinZip を起動します
2. Winzip の [Open] ボタンをクリックし、Open Archive ウィンドウを開きます。Winmorph.zip を指定し、ファイルを開きます
3. WinZip の [Extract] ボタンをクリックし、[Extract] ウィンドウを開きます
4. [Extract To] に、希望する展開のためのディレクトリ名を入力します。例えば、C ドライブの morph ディレクトリに展開するときには \C:¥morph" と入力します。Winmorph の展開が正しく実施された場合、\C:¥morph" フォルダに \AsycFilt.dil", \GAPI32.dil", \winmorph.exe" などの 23 個のファイルが作成されます。
5. WinZip を終了します

2.1.3 Winmorph のセットアップ

1. Winmorph を展開したフォルダ (C:¥morph) から \Setup.exe" をダブルクリックし、[OK] をクリックしてセットアップウィンドウを開きます
2. [Win Morph セットアップ] ウィンドウ内の左上にあるボタンをクリックして、セットアップを開始します
3. セットアップが完了します

2.1.4 Breakfast のダウンロードと展開

Breakfast のダウンロードと展開、およびインストールの仕方については、HTML 形式のオンラインドキュメントを参照してください。

< Breakfast のダウンロード >

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

2.1.5 wjpmph15 のダウンロードと展開

wjpmph15 は、IFS 汎用日本語形態素解析規則 のことです。IFS 汎用日本語形態素解析規則は、東京外国語大学で開発された日本語形態素解析規則です。ICOT フリーソフトウェアの一つである形態素解析システム LAX の解析規則部分を JUMAN 2.0(奈良先端科学技術大学院大学・松本研究室からリリースされている形態素解析システム)の形態素解析エンジン用に移植し改良したもので、新たに辞書データを含んでいます。この形態素解析規則は、日本語サブセットを分節(形態素解析)する能力を持っています。この形態素解析規則が前提とする形態組織の体系は、いわゆる構文文法の品詞体系にできるだけ依存しないものであり、そのために、この規則による解析の結果は、中立性が高く言語研究から各種アプリケーションまで応用分野が広く応用できることが特徴です。

wjpmph15(.tgz) は 下記のサイトからダウンロードします。

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

このファイルは圧縮されていますので、Nico Mak Computing, Inc. の WinZip (version 6.1 以降) などを使って展開します (WinZip を使った展開の仕方は、上記を参照してください)。うまく展開されると、\dic", \dict", \rule", \README.txt" などのフォルダやファイルが作成されます。

2.1.6 wjpmph15 の \dic" のコピー

winmorph.exe のあるディレクトリに、4.1 で示した方法で展開した wjpmph15 内のフォルダ \dic" をコピーし、C:¥morph¥dic を作成します。

エクスプローラから **wipmph15** を開き、**\dic"** をそのまま右ドラッグし、マウスのポインタを **Winmorph** ディレクトリに合わせます。手を放し、**[ここにコピー]** をクリックして、**\dic"** を ドロップします。

2.2 Winmorph のソースコードの修正について

この節では、**Winmorph** のソフトウェアコードを修正するためのプラットフォーム環境の整備について説明します。

Winmorph は **Microsoft (R) VisualBasic (R) 5.0** 版で開発されています。そのため次の手順に従って、開発環境を整えてください。

1. VB5.0 のセットアップ

Winmorph のアプリケーションを使用するために、**Microsoft Visial Basic 5.0** 版「**VB5.0**」をセットアップします。インストール時に、標準セットアップを選択してください。

2. VBSP3 のダウンロードとセットアップ

VBSP3 を **Microsoft** 社のホームページにあるソフトウェアサービスサイトからダウンロードします。自動解凍形式なので、展開後「**setup.exe**」を実行しセットアップを行います。展開すると **all**, **chs**, **deu**, **seup.inf**, **setup.exe** などのフォルダやファイルを含む **\Vs97.sp3** フォルダが作成されます。正しくインストールすると、**VB5.0** を起動したときに「**SP3**」の表示が初期画面に現れます。

< **Microsoft** 社 ホームページ >

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

3. Winmorph (ソースコード) のダウンロードと展開

Winmorph は **ZIP** 圧縮形式で提供されます。**WinZIP** を使って展開します。

3 基本操作

第3章では、Winmorph の基本的な操作方法について説明します。

- è Winmorph の起動
- è Winmorph のウィンドウ構成
- è Winmorph のインタフェースの利用方法
- è Winmorph の終了

3.1 Winmorph とは

Winmorph とは、形態素解析エンジン (Breakfast(R)) を利用して形態素解析規則と形態素辞書を作成するためのワークベンチです。新しく形態素解析規則を作成するだけでなく、既存の形態素解析規則を修正したり、必要に応じて形態素辞書を一部削除したり、新規追加をすることが可能です。

次の Winmorph の機能を使用して、容易に形態素解析規則や辞書を作成したり変更したりすることができます。

- è 形態素解析規則や形態素辞書の作成
- è 既存の形態素規則や形態素辞書の修正と追加
- è 形態素解析規則と形態素辞書のパッケージ化とそれらを選択的に利用した解析実験

3.2 Winmorph の起動

本章の説明では、Winmorph と Breakfast があなたがお使いのプラットフォームにインストールされていることが前提です。

< Breakfast のダウンロード >

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

エクスプローラーから、C:¥morph¥winmorph.exe ファイルを指定して、ダブルクリックします。

3.2.1 その他の Windows プラットホームからの起動

[スタート]メニューの[ファイル名を指定して実行]を選択し、\C:¥morph¥winmorph.exe" と入力します。[Enter]を押して起動します。

ì 註: _____	è
winmorph.exe のパスは、\Winmorph" の展開のときに指定したものです	
í _____	è

3.3 Winmorph の概要

Winmorph を起動すると図1で示すような Winmorph ウィンドウが表示されます。

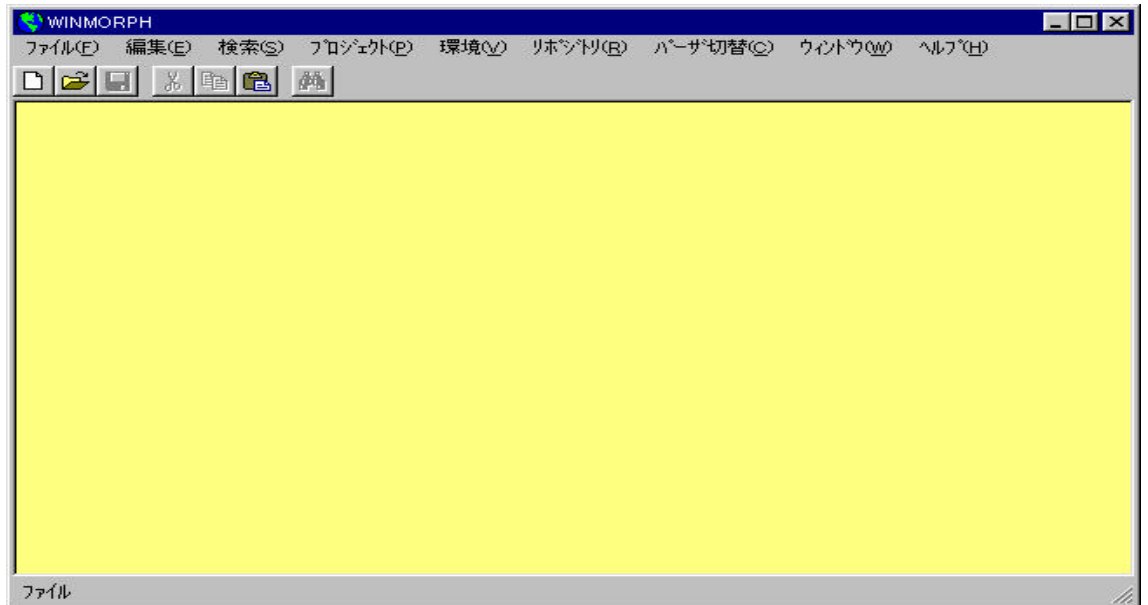


図 1: 起動後の表示ウィンドウ

3.4 Winmorph ウィンドウ

Winmorph ウィンドウを図2に示します。図2に示すように、ウィンドウは「タイトルバー」「メニューバー」「ツールバー」「入出力作業領域」の4つの領域があります。










図 2: Winmorph ウィンドウ

3.4.1 ウィンドウメニュー

メニューバーにあるメニューから **Winmorph** の各種機能にアクセスすることができます。メニューの要約を示します。

3.4.2 ツールバー

メニューバーが提供する各種機能のうち煩雑に利用する機能についてはツールバーからアクセスできます。これらのボタンの概要を示します。

ボタン	コマンド
	新規作成。テキストと解析結果の2つのウィンドウが開かれます
	開く。既存のファイルを開きます
	名前をつけて保存。ファイルを保存できます
	切り取り。選択したテキストを切り取ります
	コピー。選択したテキストをコピーします
	貼付け。カット、コピーしたテキストを貼り付けます
	検索。指定した文字列の検索をします

3.4.3 解析実行と結果表示ウィンドウ

入出力作業領域 形態素解析の基本機能は、文字入力に対して形態素解析を実施しその結果を表示するものです。入出力作業領域は、文字入力のためのウィンドウと解析結果出力がされるウィンドウが表示される領域です。

文字入力のためのウィンドウは、他のアプリケーションからドラッグアンドドロップを使って文章を貼り付けることができます。結果出力のためのウィンドウからは、ドラッグアンドドロップを使って他のアプリケーションに結果を貼り付けることができます。

3.5 Winmorph の終了

Winmorph を終了するには、Winmorph ウィンドウで [ファイル] メニューから [終了] を選択します。

メニュー	サブメニュー	コマンド
ファイル	新規作成	ファイルの新規作成を行います。通常は、テキストと解析結果の2つのウィンドウを開きます
	開く	[ファイルを開く]のダイアログボックスを表示し、指定したファイルを開きます。これは、エクスプローラやファイルマネージャからのドラッグアンドドロップでも可能です
	上書き保存	編集中のファイルを同じファイル名で上書き保存します
	名前をつけて保存	Oを付けて保存ダイアログボックスを表示し、指定したファイル名でファイルを保存します
	印刷	ダイアログボックスを表示し、テキストを印刷します
エディタを閉じる	winmorph の終了	winmorph を終了し、現在編集中のテキストには保存するかしないか尋ねます
編集	元に戻す	1つ前の編集操作を取り消して元に戻します
	切り取り	選択したテキストを切り取り、クリップボードにコピーします
	コピー	選択したテキストをクリップボードにコピーします
	貼り付け	カーソル位置にクリップボードの内容を貼り付けます
	削除	選択したテキストを削除します
	すべて選択	編集中のテキストの全範囲を選択します
検索	検索	[文字列の検索]ダイアログボックスを表示し、指定した文字列を検索します
	次を検索	上記に同じ
プロジェクト	プロジェクト定義	[プロジェクト定義]ダイアログボックスを表示し、規則・辞書ファイルを決定できます

メニュー	サブメニュー	コマンド
検索	パーザ作成	[パーザの作成] ダイアログボックスを表示し、指定された規則・辞書ファイルからパーザを作成します
	パーザ実行	テキストウィンドウのテキストをパーザを用いて解析します
環境	環境変数設定	Breakfast のパス設定や解析の表示方法について、環境を設定します
	ツールバー表示	チェックするとツールバーを表示します
	ステータス表示	現在の編集の状態を winmorph ウィンドウの下に表示します
リポジトリ	規則ファイル登録	[規則ファイルの選択] ダイアログボックスを表示し、規則ファイルを指定できます
	辞書ファイル登録	[辞書ファイルインデックスの登録] ダイアログボックスを表示し、辞書ファイルを指定できます
	リソースファイル登録	リソースファイルを指定します
パーザ切替		chasen か breakfast にパーザを切り替えます
ウィンドウ		テキストウィンドウと解析結果ウィンドウをこの画面から選択できます
ヘルプ		オンラインドキュメントとガイドを開きます

4 形態素解析

第4章では、Winmorph を使った形態素解析の実際について説明します。

è winmorph から breakfast の形態素解析機能をインターラクティブに利用する

è wjpmph15 を使って、形態素解析ソフトウェアを作成し、その後形態素解析を行う

ì 註 _____ è
| winmorph で 形態素解析を実行する前に、形態素解析エンジン (breakfast) と形態
| 素解析規則と形態素辞書 (wjpmph15) がインストール済みであることを確認して
| ください。 |
í _____ è

4.1 形態素解析の手順

形態素解析の実行には、

1. すでにコンパイルされた形態素解析パーザを使って実行する場合
2. 形態素解析規則と形態素辞書を使って、形態素解析パーザを作成してから実行する場合

があります。

4.2 準備

メニューバーにある[環境変数設定(V)] で Breakfast の実行バイナリ (\cbf.exe") があるパスを入力します
例えば、です。

4.3 コンパイルされた形態素解析パーザを使う

この節では、コンパイルされた形態素解析パーザを使って Winmorph を利用し形態素解析を行う方法を示します。

1. \winmorph.exe" をクリックし、Winmorph を起動します
2. メニューバーの [ファイル(F)] から [新規作成(N)] を選び、入出力作業領域に新規に2つの作業ウィンドウを開きます。上側が入力作業ウィンドウです。下側が結果出力ウィンドウです
3. 入力作業ウィンドウに日本語のテキストを入力します。もしくは、他のアプリケーションからドラッグアンドドロップで文章を貼付けます

4. メニューバーの [プロジェクト (P)] から [パーザの実行 (P)] を選びます。形態素解析が実施され、解析の結果が結果出力ウィンドウに表示されます
5. 出力結果を保存する場合、保存したいデータが表示されているウィンドウ (解析結果を保存する時は、結果出力ウィンドウ) をクリックし、メニューバーの [ファイル (F)] をクリックし、メニューを出します。メニューの [上書き (S)] または [名前をつけて保存 (A)] を選びます。ウィンドウはマウスで直接選ぶか、メニューの ウィンドウ (W) から選びます

} *
é

{
è

入力作業ウィンドウの替わりに、予め作成したテキストファイルを用いる場合、[1] の手順でメニューバーの [ファイル (F)] から [開く (O)] を選びます。2つのウィンドウのうち、タイトルバーにそのファイルのパス名をもつウィンドウが上に現れます

4.4 形態素解析規則ファイルと辞書ファイルからパーザをコンパイルする

この節では、形態素解析規則と形態素辞書を使ったパーザをコンパイルし、その形態素解析パーザを使って **Winmorph** を利用し形態素解析を行う方法を示します。

1. "\winmorph.exe" をクリックし、**Winmorph** を起動します
2. メニューバーの [リポジトリ (R)] から [規則ファイル登録 (R)] をクリックし、規則ファイルを登録します。
3. rc ファイル、品詞分類辞書などの各ファイルのパスを指定します。[参照] をクリックして 任意の規則ファイルを選択することができます。例えば、**wjpmph15** の場合、rc ファイルの参照として
dic から **jumanrc** を選択してファイルを開くと、これに対応した辞書が全て各パス名とともに自動的に割当てられ、ディスプレイに表示されます。各ファイルはユーザーが指定し、[登録] をクリックします
4. 登録した規則ファイルに名前をつけ、[登録] をクリックします "\(%le name).rif" のファイルが新たに
dic の中に作成されます
5. メニューバーの [リポジトリ (R)] から [辞書ファイル登録 (D)] をクリックし、辞書ファイルを登録します。規則ファイルの指定と同様に、参照したい辞書のパスを指定します。辞書ディレクトリ変更 をクリックすると参照して辞書ファイルを開くことができます。このとき
dic を指定すると、その中の全ての辞書ファイルが画面左に現れ、ここから ファイルを選んで 追加 をクリックするか、単に 全追加をクリックします。画面の

右に選んだファイルが表示されます。訂正は、削除、全削除 をクリックして行います

6. 登録した辞書ファイルに名前をつけ、登録 をクリックします。\"(file name).dif\" のファイルが新たに dic の中に作成されます

7. メニューバーの [プロジェクト (P)] から [プロジェクト定義 (p)] を選び、[規則ファイルディレクトリ/規則インデックスファイル] のパスを指定します ([参照] から作成したインデックスファイルを指定できます。通常は、jumanrc などがあるディレクトリ dc を指定します)。[辞書ファイル/辞書インデックスファイル] については、[追加]、[削除] で操作します (作成したインデックスファイルを指定できます)。設定を終えたら、[OK] をクリックします

8. メニューバーの [プロジェクト (P)] から [パーザ作成 (C)] を選び、[実行] をクリックして、登録した規則ファイルと辞書ファイルのコンパイルを行います。

註	_____	è
	コンパイルにはかなりの時間を要します	
í	_____	è

9. コンパイルが終了すると、形態素解析ソフトウェアが作成されます。次に、前節で説明した「コンパイルされた形態素解析パーザを使う」の手順に従って、形態素解析を行います。

5 形態素解析の機能概念

第5章では、形態素解析と形態素解析ソフトウェアの機能概念を説明します。

- è 形態素解析と形態素解析ソフトウェア
- è 動作概念説明
- è 機能概念説明
- è 形態素解析駆動データ部とそのデータ作成

5.1 形態素解析

形態素 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します。この意味を有する最小言語形式を「形態素 (morpheme)」と言います。以下、形態素の言語的な説明を森岡健二「日本文法体系論」から引用して説明します。形態素は形式の点から2種類に分けられます。(1) 他の形式から切り離して単独で発音されるものを自立形式 (free form) と呼んでいます。(2) 必ず他と結合して一続きに発音されるものを結合形式 (Bound form) と呼んでいます。「単語を分割してゆくと」を例文とすると、「単語を」「単語」「分割してゆくと」「分割して」などが自立形式で、「を」「してゆく」「ゆくと」「と」などが結合形式です。

語基、派生辞、屈折辞

派生辞 とは派生したり屈折したりする形式の基幹となる形態素で、上記の例では「単語」「分割す(る)」などです。

語基 は、語基に結合して派生語を作る形態素です。上記例の「(し)てゆ(く)(と)」や、接頭辞や接尾辞がこれに相当します。

屈折辞 は、語基に結合して屈折語を作る形態素です。「(単語)を」「(分割し)て」「(ゆ)く」等、助詞や活用語尾が相当します。

例えば、「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。」という文では、「単語」「を」「分割す(る)」「ゆ(く)」「と」「それ」「以上」「分割でき(る)」「な(い)」「最小」「の」「形式」「に」「達す(る)」がそれぞれ形態素です。これら形態素を一覧にして表に示します。

形態素	形態素の分類
「単語」	語基
「を」	屈折辞
「分割す(る)」	語基
「て」	派生辞
「ゆ(く)」	派生辞
「く」	屈折辞
「と」	屈折辞
「それ」	語基
「以上」	語基
「分割でき(る)」	語基
「き」	派生辞
「な(い)」	語基
「い」	屈折辞
「最小」	語基
「の」	屈折辞
「形式」	語基
「に」	屈折辞
「達す(る)」	語基
「る」	屈折辞
「。」	語基

日本語の文は特徴として形態素の切れ目がなく、いろいろな形態素が直接つながっています。与えられた文を形態素に分割し、一つ一つの形態素を辞書を参照しながら分類名と対応付ける操作を形態素解析 (Morphological Analysis) と言います。

5.2 形態素解析用ソフトウェア

形態素解析ソフトウェアは、コンピュータを使って形態素解析を行うソフトウェアです。入力された文字列を区切り、形態素辞書に登録されている辞書項目を参照しながら文を形態素の連鎖に分割します。形態素の区切り方の妥当性を記述したものは形態素解析規則と呼ばれます。同じ語基でも「単語」には「を」が後続しますが、「分割す(る)」には「を」は後続しません。必要に応じて形態素の基本分類も詳細化されることがあります。このように形態素解析は、

1. 文字を区切ったり、その区切り方を制御する部分と
2. 区切られた文字列が形態素辞書に登録されているかどうかを判断する部分
3. 区切った形態素どうしの接続の妥当性を調べる部分

から構成されます。

1 を形態素解析エンジン、2 と 3 をあわせたものを形態素解析駆動データと言います。

5.3 動作概念説明

形態素解析ソフトウェアのソフトウェアとしての働きは、文字列を入力すると、文字列の形態素連鎖を結果として出力する、というものです。ソフトウェアの動作概念を次の図で示します。

形態素解析エンジンと形態素解析駆動データをあわせて形態素解析器 (parser) と言います。

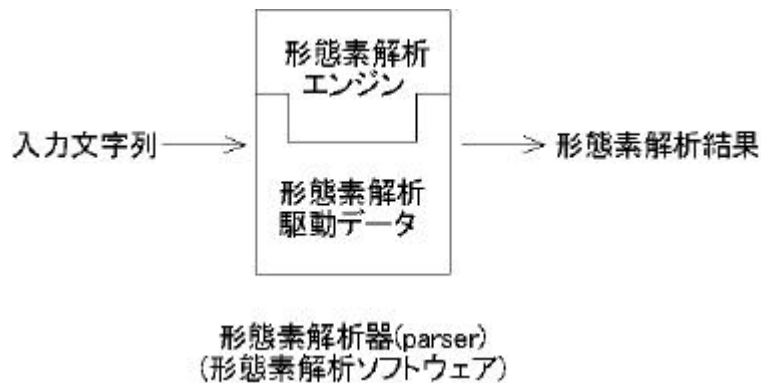


図 3: 動作概念図

5.4 機能概念説明

形態素解析ソフトウェアの基本機能は入力文字列を形態素の連鎖に変換することです。例えば、次のような入力文を解析したとします。

入力文字列 > 「日本語の形態素解析の機能を説明します」

その形態素解析結果は次のようになります。

形態素解析結果 > 「-日本語-の-形態素-解析-の-機能-を-説明-し-ま-す-」

付随情報として形態素の分類名や読み等の形態素辞書データが得られます。

形態素	形態素分類名	読み
「	開き括弧	ひらきかっこ
日本語	体言語基	にほんご
の	連体助辞	の
形態素	体言語基	けいたいそ
解析	体言語基	かいせき
の	連体助辞	の
機能	体言語基	きのう
を	格助辞	を
説明	体言語基	せつめい
し	用言語基	「する」し
ま	丁寧接辞	「ます」ま
す	活用助辞	「す」す
」	閉じ括弧	とじかっこ

5.4.1 形態素解析駆動データ部

形態素解析駆動データ部には、形態素解析エンジンが文字列を形態素連鎖に変換するために利用するデータ(形態素辞書データと形態素接続規則)が記録されています。

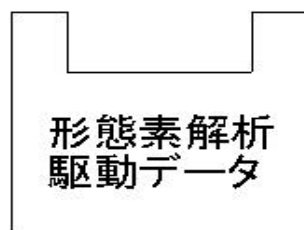


図 4: 駆動データ

形態素解析ソフトウェアを作成する場合、この2つのデータは一般には人手で作成します。国語学や言語学の知見を利用し、人間が理解できる形式で、データファイルとして作成します。つまり、形態素辞書データを記述したファイルと形態素接続規則を記述したファイルを作成することになります。そして、そのデータファイルを専用の変換ソフトウェアを使って形態素解析エンジンで利用可能なデータ形式にします。このようにして形態素解析駆動データが出来上がり、この駆動データと形態素解析エンジンを組み合わせることで形態素解析を行うことができます。

形態素解析ソフトウェアのメイン機能部分(形態素解析器(parser))の構成法を次に示します。

- [1] テキストエディタを使ってファイルを作成します
- [2] 作成したファイルのデータ形式の変換を行います
- [3] 形態素解析器(parser)を動作状態にします
- [4] 入力文を与え形態素解析します

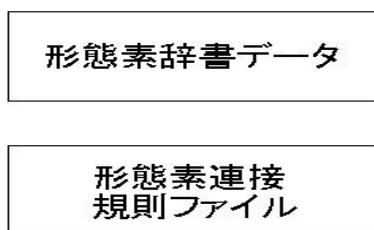


図 5: ファイル

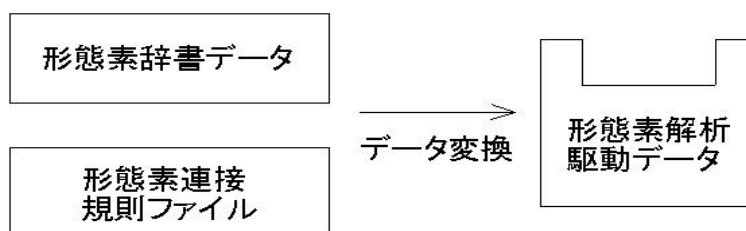


図 6: データ変換

5.5 データの作成

5.5.1 形態素辞書データ

形態素辞書データの作成は、それぞれの形態素についてその形態素に固有の項目を記述してゆくことです。形態素辞書の基本な項目は、形態素エントリーの形(見出し形式)、形態素分類名、読み、その他の情報です。それら項目を構造的に記述したものが以下です。

⌈	形態素表記(エントリー)	_____	è
	(a) 形態素表記(辞書エントリー) (b) 形態素分類名 1(形態素分類名 2, 形態素分類名 3,...) (c) 読み (c) その他の辞書情報		
⌋	_____		è

例を挙げます。

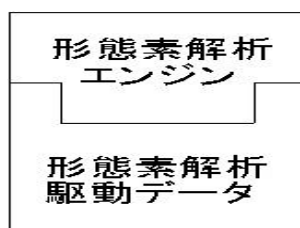


図 7: 形態素解析器

\単語"	形態素表記:	単語
	形態素分類名:	体言語基
	読み:	たんご
	その他の辞書情報:	-
\分ける"	形態素表記:	分け
	形態素分類名:	用言語基, ウ系弱変化
	読み:	わけ
	その他の辞書情報:	-

具体的な書き方の書式は、形態素解析ソフトウェアによって異なります。

5.5.2 形態素連接規則の作成

形態素どうしの結合の妥当性(どの形態素にどの形態素が接続するのか)を記述したデータが形態素連接規則です。

連接規則の記述の基本 形態素を取り上げて、その形態素を中心としてどのような形態素に結合する可能性があるのかを網羅的に記述します。例えば、次のような概念図を考えることができます。

```

形態素A  接続  形態素B
形態素A  接続  形態素C
...

```

別の方法では、次に示すような接続マトリックスとしても記述可能です。

	形態素A	形態素B	形態素C	形態素D	...
形態素A	x	x	x	x	
形態素B		x		x	
形態素C			...		
形態素D	x	x	...		
...					

具体的なデータの書き方の書式は、ソフトウェアによって異なります。