形態素解析視覚化システム ViJUMAN

美 寿 満
び じゅ まん

Version 1.0 使用説明書

山下 達雄　　松本 裕治

平成8年2月9日

Visual Interface for Morphological Analysis System **ViJUMAN** Manual

Version 0.1	11 October 1995
Version 0.9	18 December 1995
Version 1.0	9 February 1996

はじめに

本システムは、「JUMANの出力結果を見やすく図示できないか」「JUMANの文法でタグ付きコーパスを作成するためのツールはないのか」「新しい単語を簡単に登録できないか」という要望に答えるために開発されました。

本システムには以下のような特徴があります。

- JUMANの出力を視覚化(図示)することにより、解析結果が分かりやすくなりました。
- 出力結果から任意に形態素列(パス)を選ぶことができ、さらに部分再解析・形態素情報の修正などの機能も実装されているので、タグ付きコーパス作成が楽になります。
- 文法辞書に基づき矛盾なく簡単に新しい単語を登録できます。

基本的に個人で適当に開発したシステムであり、さまざまな不備があると思います。今後も可能な限り改良を重ねるつもりですので、何かありましたら御指摘下さい。

本システムは奈良先端科学技術大学院大学松本研究室の方々の御協力を得て完成しました。山田篤先生には解析結果のより詳細な情報を得るためにJUMANを改良して頂きました。御協力頂いた各位にお礼申し上げます。

平成7年12月　山下 達雄

本システムに関するお問い合わせは以下にお願いします。

〒630-01
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(松本研究室)
Vi JUMAN管理担当者
Fax: (07437)2-5249
Email: vijuman-adm@cactus.aist-nara.ac.jp

また以下のURLにて最新情報を提供しています。

URL: http://cactus.aist-nara.ac.jp/student/tatuo-y/vijuman.html

目 次

図目次

1 このマニュアルの読み方

このマニュアルは、日本語形態素解析システム JUMAN についてある程度の知識を前提として作成されている。JUMAN について全く何も知らないならば、まず文献 [1] を読むことをお勧めする。

美寿満のインストール方法については、インストールガイドが別に用意されているのでそちらを参照されたい。

また、すぐにでも使いたいという人は第 3章 (p.3) にチュートリアルが用意されているので、そこから読み始めるとよい。

表記について

< >　システム上のボタンまたはサブメニュー上のボタンを表す。例 : <解析する>
　▼ はメニューボタン (押すとメニューが現れる) を表す。例 : <ヘルプ▼ >
「」　本システムで導入された特別な概念を表す。原則として初出のときのみ。例 : 「ダミー」
　ただし分かりにくいと思われるものについてはこの限りではない。

2 システムの概要

2.1 目的

今までの JUMAN [1] の解析結果情報は 1 次元に出力され、曖昧性がある場合、どの形態素とどの形態素が接続しているかが分かりにくかった。『今日もしないとね。』を `juman -m -e` で解析した結果を整形して図 1に示す。

```
今日    きょう    今日    名詞    6    時相名詞            9    *                0    *           0
今日    こんにち  今日    名詞    6    時相名詞            9    *                0    *           0
も      も        も      助詞    9    副助詞              2    *                0    *           0
もし    もし      もし    副詞    8    陳述副詞            6    *                0    *           0
し      し        する    動詞    2    *                   0    サ変動詞         16   基本連用形  8
しない  しない    しなう  動詞    2    *                   0    子音動詞ワ行     12   基本連用形  7
しない  しない    しない  名詞    6    普通名詞            1    *                0    *           0
ない    ない      ない    形容詞  3    *                   0    イ形容詞アウオ段 18   基本形      2
ない    ない      ない    接尾辞  14   形容詞性述語接尾辞  4    イ形容詞アウオ段 18   基本形      2
と      と        と      助詞    9    格助詞              1    *                0    *           0
と      と        と      助詞    9    名詞接続助詞        4    *                0    *           0
と      と        と      助詞    9    述語接続助詞        5    *                0    *           0
と      と        と      助詞    9    引用助詞            3    *                0    *           0
ね      ね        ね      助詞    9    終助詞              6    *                0    *           0
。      。        。      特殊    1    句点                1    *                0    *           0
```

図 1: 今までの JUMAN の出力

そこで Tcl/Tk [3][4][5] によってこの解析結果を束状[1] に図示することによって解析結果を認識しやすくした (図 2)。この図を「束図」と呼ぶことにする。

またこれに、任意のパスを選ぶための機能や形態素情報を修正して辞書登録する機能等を付け加えた。これによりタグつきコーパス作成支援システムとして利用できるようになった。

---

[1] 数学的な束 (latice) とは異なる。

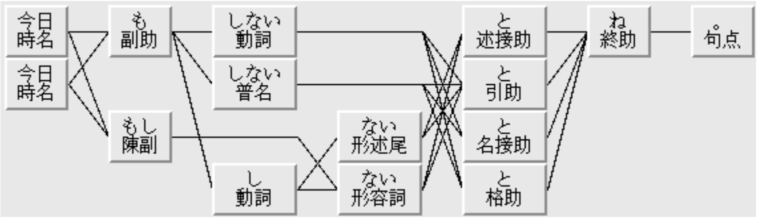


図 2: JUMAN の解析結果の視覚化

2.2 概観

システムの概観を図 3 に示す。

形態素解析は Prolog 版 JUMAN 用のサーバプログラム (`juman_pl`) [2]を用いている。その理由は以下の通り。

- 既存の C 版 JUMAN の出力には必要な情報 (識別子、開始位置、終了位置、前接形態素の識別子のリスト) が不足している。
- 既存の JUMAN 体系を変更したくなかった。

しかし `juman_pl` は、

- 連接コスト ・品詞コスト等の出力をサポートしていない。
- 品詞名 ・未定義語の扱いが、独特 (C 版と異なる) である。

という問題があったので、その点を改良して用いている。

Prolog 版クライアントの代わりに Perl [2] で JUMAN のクライアント (`clsfy.perl`) を作成しそれを用いている。Perl 版クライアントは`juman_pl` から解析結果を受けとり、更に表示のためにクラス分けを行なう。

Tcl/Tk で書かれたメインプログラム (`vijuman`) がこの Perl 版クライアントを呼びだし文を解析させ束図を描く。

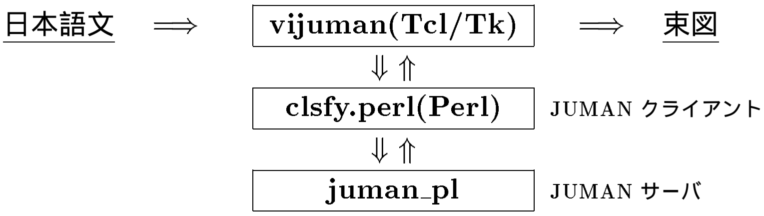


図 3: システムの概観

---

[2] 詳しくは文献 [1] を参照。

3　入門編 ~チュートリアル~

3.1　各部の名称

各部の名称について説明する (図 4を参照)。

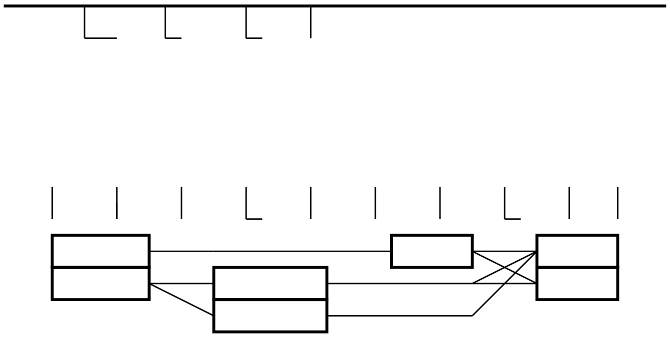


図 4: 各部の名称

入力パネル　解析しようとする文を入力したり、解析を行なうためのパネル。

- 入力文の表示と直接入力に利用される「入力スペース」
- 入力文のファイルからの読み込みやJUMANによる解析を行なう「入力メニュー」

から構成される。

情報パネル　個々の形態素情報と、JUMANでの解析に用いるコスト幅を表示する場所。

出力メニュー　最適パス・新単語データなどを出力するためのメニュー。

操作メニュー　束図の表示に関する操作を行なうためのメニュー。

キャンバス　束図を表示するためのスペース。束図自体は、

- 形態素を表す「形態素ボックス」
- 形態素どうしの接続を表す「辺」
- 表示上の都合により導入された「ダミーボックス」(実際は見えない)

から構成される。束図の構成に関しては 4.1.1(p.10) で詳しく説明する。

3.2　実際に動かす

ここでは実際にシステムを動かしながら基本的な操作を覚えてもらう。

3.2.1 システムの起動

美寿満システムを起動するには `vijuman` と打ち込めばよい (コマンドラインオプションについては 4.5.3(p.21) を参照)。

初めての場合リソースファイル (.vijumanrc) を作るかどうか聞かれるので < OK >ボタンを押す。

するとシステムが立ち上がるので、< ヘルプ▼ >メニューの下の< サンプル束図 >を選んでチュートリアル用の束図を表示させる。図 5のようになったであろうか ?

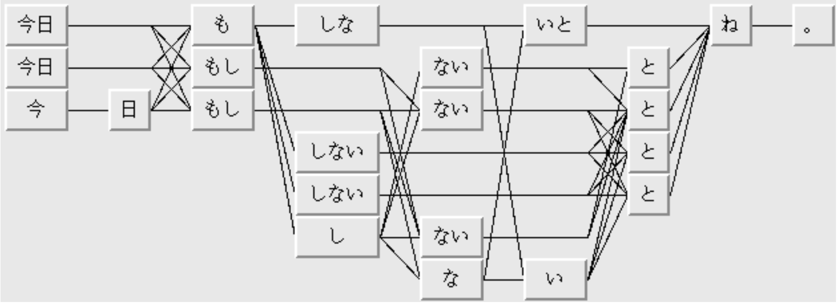


図 5: チュートリアル用サンプル束図

3.2.2 情報パネル

しかし、この図だけでは見出し語 (今日 , もし など) しか分からない。もっと詳しい情報が得たいときはマウスカーソルを目的の形態素ボックスにあわせる。すると、図 6のような詳しい情報が情報パネルに表示される (し の場合)。

見出し語 し ■
読み し □
基本形 する □
意味 0 □

形態品詞 動詞 □
品詞細分類 * □
活用型 サ変動詞 □
活用形 基本連用形 □

最適コスト 329 □
品詞コスト 100 □
本当のコスト ?

図 6: 形態素の詳細情報

3.2.3 形態素ボックス上の情報

いちいちマウスカーソルをあわせて個別に情報を見るよりも、全ての形態素ボックス上に情報が表示してあったほうが、全体を見るには便利である。

情報パネル上のチェックボタンにより形態素ボックス上に表示する要素が選べる。選んだ後に< 描き直し > (マウスの 3 (左) ボタン) が必要。図 6では「見出し語」しか選択されていないが、図 7のようにさまざまな情報が表示できる[3]。

---

[3] 図中の「時名」は時相名詞の省略形。詳しくは 4.1.3(p.11) を参照。

見出し語	見出し語, 形態品詞	最適コスト, 品詞コスト
今日	今日 時名	139 99
何も表示しない	見出し語, 形態品詞, 読み	見出し語, 形態品詞, 品詞細分類, 品詞コスト
	今日 時名 きょう	今日 名詞 時相名詞 99

図 7: 形態素ボックス上の情報の例

3.2.4 最適パスの強調

さて次に最適パス[4]を調べたいとする。マウスの 1 (左) ボタンで形態素ボックスをクリックするとその形態素を通る最適パスが盛り上がって紫色で強調される。このパス上の形態素ボックスはシステム内部に記憶され、出力のときに用いられる。

<お色直し>をするか、形態素ボックス以外のところでマウスの 2 (中) ボタンをクリックすると強調がなくなり最適パスはキャンセルされる。

3.2.5 形態素ボックスの削除

さてサンプルの束図を見てみると もし (動詞『もす』(模す・燃す?) の連用形) や しない (市内・竹刀?) といった余計なものがいっぱいある。このような形態素が入らないようにパスを選択するには、事前にこれらの形態素ボックスを削除してしまえば良い。

<削除モード>メニューをクリックすると削除モードに入る。もう一度クリックすると通常モードに戻る。削除モードの状態でマウスの 1 (左) ボタンで形態素ボックスをクリックするとその形態素ボックスが削除され、さらにその影響で浮いてしまう形態素ボックスも削除される (詳しくは 4.2.5(p.14) を参照)。

図 8(a) は削除モードで も をクリックした結果である。 も 以外にも複数の形態素ボックス (し , しな , しない など) が結果として削除されている。

削除にはもう 1 つ別の方法がある。形態素ボックスをマウスの 2 (中) ボタンでクリックすると、クリックした形態素ボックスにどうやってもたどり着けない (接続できない) 形態素ボックスがすべて削除される (詳しくは 4.2.6(p.14) を参照)。同時に最適パスも強調される。

図 8(b) は図 5の状態 (図 8(a) の状態から<削除全復活>を押せばよい) で しな をマウスの 2 (中) ボタンでクリックした結果である。 もし , ない , しない 等は しな に接続できないので削除されている。

削除されたボックスは黒色になるが、マウスの 3 (右) ボタンを (画面上の任意の場所で) 押すことによって画面上から消える。図 8(c) は図 8(b) の状態ででマウスの 3 (右) ボタンを押した結

---

[4] 最適パスとは、それぞれに一番尤もらしい品詞情報の付加された形態素の並び。例えば『大人だ』という文は、『大人 (おとな:名詞) だ (判定詞)』や『大 (だい:接頭辞) 人 (ひと:名詞) だ (判定詞)』というふうに解析されるが、一番尤もらしいのは前者であり最適パスといえる。尤もらしさの判定にコストが用いられる。

果である。

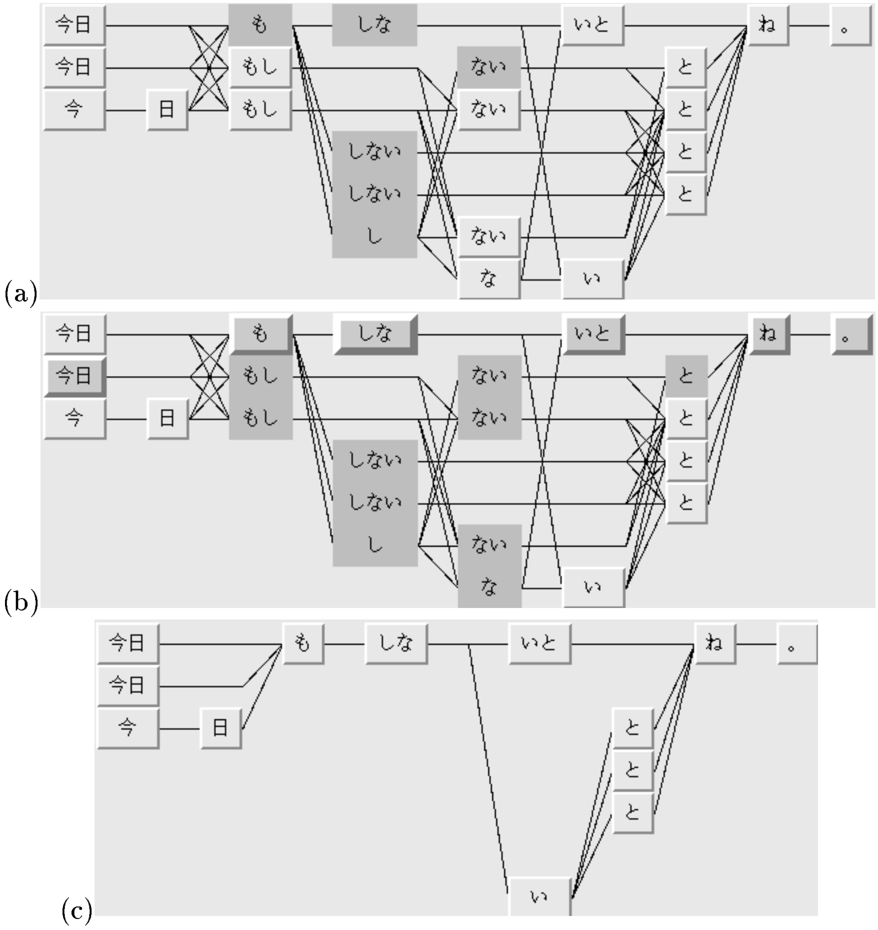


図 8: 形態素ボックスの削除の実行例

3.2.6 削除のキャンセル

削除された形態素ボックスを復活させる機能 (＜削除 UNDO ＞ ＜削除全復活＞) もある。図 8 の (a)(b)(c) のどの状態でも＜削除全復活＞ボタンをクリックすると、図 5の状態に戻る。

3.2.7 練習問題 【1】

今までの機能で、サンプルの束図から
今日 (こんにち) − もし (動詞) − な − いと − ね − 。
というパスを選択せよ

解答例 (1)

まず削除モードで全ての と と も を削除する。それから削除モードを解除して 今日 (こんにち) を 1 ボタンでクリックする。

解答例 (2)

今日 (こんにち) , もし (動詞) をそれぞれ2ボタンでクリックする。

3.2.8 パスの出力

このようにして選んだパスをタグ付きコーパスのデータとしてファイルに出力できる。

さてここで出力パネルの＜ファイル出力＞ボタンを押すと紫色で強調されている最適パス上の形態素が、確認のためにウィンドウに表示された後、指定されたファイル (デフォルトでは output_file) に出力される。

以下に練習問題で選んだパスの出力を示す。

```
今日 こんにち 今日 名詞 時相名詞 * *
もし もし もす 動詞 * 子音動詞サ行 基本連用形
な な だ 判定詞 * 判定詞 ダ列基本連体形
いと いと いと 名詞 固有名詞 * *
ね ね ね 助詞 終助詞 * *
。 。 。 特殊 句点 * *
EOS
```

UNIX のコマンドラインからも (cat や less で) ファイル output_file を見てみよう。このようにして解析済データを蓄積していくことができる。

出力のフォーマットはデフォルトのままだが自由に変更できる。5.1.3(p.23) を参照。

3.2.9 文の入力と解析

解析したい文は、入力パネル上の入力スペースに文を直接入力したり[5]、日本語文の入っているファイルから1文ずつ入力することができる。ファイルから入力する場合は＜入力ファイルの選択＞ボタンで入力ファイルを選び＜1文読み込み＞ボタンで一行ずつ読み込む。

入力メニューの＜解析する＞ボタンをクリックすることにより、JUMAN で解析され結果が束図表示される[6]。

さっそくやってみよう。＜ヘルプ▼＞の下の＜サンプル文＞を選ぶと入力スペースに以下の文が表示される。

これは美寿満です。

＜解析する＞ボタンで解析してみよう。コスト幅や文法に依存するがだいたい図9のようになるはずである (表示情報として「見出し語」と「読み」を選択してある)。

3.2.10 部分無解析

『美』『寿』『満』とバラバラになっている単語を『美寿満』という単語にまとめたい。

このように文字列中の任意の部分文字列を1つの単語にする機能が部分無解析である。

まず入力スペースで『美寿満』の部分をマウスでセレクト (マウスボタンを押しながらマウスを動かす) する (図10)。

それから入力メニュー上の＜部分無解析＞を押す。すると束図は以下のようになる。

---

[5] ローマ字かな変換機能により、かな入力ができる (4.5.1(p.20) を参照)。ただし、漢字変換はできない。マウスのカット&ペースト機能を用いると良い。

[6] ただし、`juman_pl` が起動されていることが前提となってる。詳しくは 4.5.2 (p.20)『JUMAN サーバの起動』を参照。

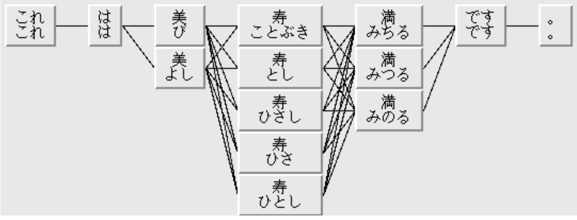


図 9: サンプル文の解析例

これは美寿満です。

図 10: マウスによるセレクト

これ ─ は ─ 美寿満 ─ です ─ 。

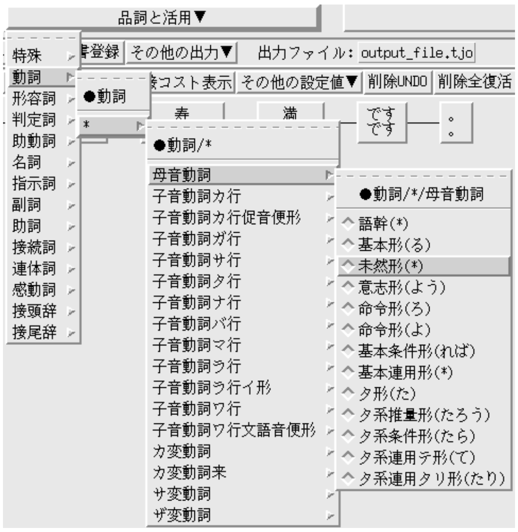
同様に、セレクトした部分を異なるコスト幅を用いて JUMAN で再解析をする機能 < 部分再解析 > もある。詳しくは 4.4.5(p.19) を参照。

3.2.11 形態素情報の修正

さて新たに作られた単語 『美寿満』は品詞 ·活用が定義されていないので修正したい。

まずマウスを 『美寿満』と表示されている形態素ボックスにのせ、情報パネルに形態素情報を表示させる。そしてそのまま情報パネルの < 品詞と活用▼ > ボタンを押して品詞 ·活用情報を修正する (図 11)。

この場合は固有名詞になる。



出力ファイル: output_file.tjo

図 11: 品詞 ·活用情報の選択的な変更

また読みが意図したものになっていないので[7]、読みの欄に直接 『びじゅまん』と打ち込んで修正する[8]。修正後はリターンキーまたは < 確定 > ボタンを忘れずに。

---

[7] ここの環境では 『びことぶきみちる』となっていた。

[8] 『bijumaN』とローマ字で打ち込み最後にシフトキーを押す。詳しくは 4.5.1(p.20) を参照。

3.2.12 自動辞書登録

＜辞書登録＞ボタンにより、この新たに定義した単語　『美寿満』を JUMAN の辞書に追加することができる。これにより次の解析からは苦しめられることはない。

自動辞書登録機能の設定については 4.3.4(p.16) を参照。

3.2.13 練習問題 【2 】

美寿満はジュマンのＧＵＩである。

美寿満で上の文を解析し[9]、形態素情報を正してファイルに出力せよ。

ここのシステムでの解答例

```
美寿満├ は├ ジュマン├ の├ ＧＵＩ├ です├ 。
```

```
美寿満 びじゅまん 美寿満 名詞 固有名詞 * *
は は は 助詞 副助詞 * *
ジュマン じゅまん ジュマン 名詞 固有名詞 * *
の の の 助詞 名詞接続助詞 * *
ＧＵＩ じーゆーあい ＧＵＩ 名詞 普通名詞 * *
である である だ 判定詞 * 判定詞 デアル列基本形
。 。 。 特殊 句点 * *
```

3.2.14 システムの終了

＜終了＞ボタンを押すか、画面上の任意の場所でシフトキーを押しながらマウスの１ (左) ボタンを押せばシステムは終了する。

3.2.15 おわりに

以上、基本的な機能のみざっと述べた。それぞれの機能についてのより詳しい解説は “応用編” を参照されたい。

---

[9] エディタ等でこの文を書き、マウスのカット&ペースト機能で直接入力し、解析する。

4　応用編 ~機能の詳細~

この章では、全機能の詳細について説明する。

4.1　束図の表示

4.1.1　表示の手順

ここではJUMANの解析結果を束図として表示する手順について説明する。

まず前処理として、JUMANの出力に含まれる全形態素を以下のルールに従ってクラス分けする。

- 同じ位置から始まる形態素は同じクラスに属する。
- ある形態素に前接する形態素は必ず今のクラスより前のクラスに属する。

例えば図1 (p.1) の出力をクラス分けすると、図12のように8つのクラスに分けられる (図中の数字はクラス番号)。

0	1	2	3	4	5	6	7
今日 (きょう) 今日 (こんにち)	も もし	し しな しない (動詞) しない (普通名詞)	な ない	いと	と (格助詞) と (名詞接続助詞) と (述語接続助詞) と (引用助詞)	ね	。

図 12: クラス分けの例

次にキャンバス上に、「形態素ボックス」を表示する場所 (「ボックスフィールド」) と「辺」を表示する場所 (「辺フィールド」) を交互に配置する。各ボックスフィールドは各クラスに対応しており、先頭クラスに対応するボックスフィールドから順番に並べる。形態素ボックス間の接続は辺フィールド上の辺で表す。図13にキャンバスの構成を示す。

クラス 0		クラス 1		クラス 2	...
ボックスフィールド	辺フィールド	ボックスフィールド	辺フィールド	ボックスフィールド	...
形態素ボックス	—— 辺 ——	形態素ボックス	—— 辺 ——	形態素ボックス	...
:	:	:	:	:	

図 13: キャンバスの構成

しかし形態素ボックスどうしの接続がいくつかのクラスをまたぐ場合があり、表示上の問題が起こる (形態素ボックスと辺が重なってしまう)。そこで「ダミー」という概念を導入した。ダミーはクラス分けの時に作られる疑似形態素である。クラス分けの時、ある形態素の前接形態素が前接クラスに存在しない時、その前接形態素と同じ情報を持つダミーという疑似形態素が前接クラス内に生成される。ダミーは表示される時形態素ボックスにはならず「ダミーボックス」と呼ばれるものになり辺が描かれる (図14)。

図 14: ダミーボックス

4.1.2 ＜表示形態▾ ＞

＜表示形態▾ ＞メニューにより形態素ボックスの配置の仕方を選択できる。配置の仕方は 3 通りある。

＜コスト順＞ 各ボックスフィールド中の形態素ボックスとダミーボックスをコストの低い順に上から配置している (ダミーボックスもコストを持っていることに注意)。コストの低いものが上に寄るので、最適パスも上に寄りやすくなる。辺の交差が複雑になるという欠点がある。

＜中寄せ＞ コスト順と同様の配置法だが各クラスごとに中央に寄せてある。

＜すっきり配置＞ 文末の形態素から配置していく。もし自分のダミーが存在していたら、すでに決定している自分のダミーボックスの位置 (Y 座標) に自分の位置をあわせる。すっきりとして見やすい。

＜すっきり中寄せ＞ すっきり配置と同様の配置法だが各クラスごとに中央に寄せてある。

コスト順とすっきり配置の違いを図 15 に示す。数字は最適コスト (括弧はダミーのコスト)、太線は最適パスを表すとする。(a) はコスト順、(b) はすっきり配置による配置である。

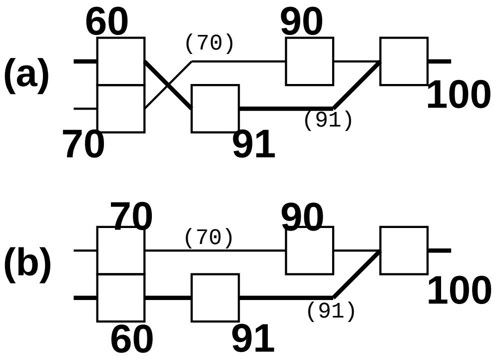


図 15: 配置法

なお表示形態を変更すると自動的に＜描き直し＞ (4.2.7を参照) が行なわれる。

4.1.3 形態素ボックス上に表示する情報の選択

情報パネル上のチェックボタンにより形態素ボックス上に表示する要素が選択できる。選んだ後に＜描き直し＞ (マウスの 3 (左) ボタン) が必要。

「形態品詞」を選択し「品詞細分類」を選択しない状態で略称表示モード (＜その他の設定値▾ ＞メニューで設定可 (4.1.6を参照)) ならば、形態素ボックス上には「形態品詞」の代わりに省略形[10] が表示される。

例えば「見出し語」「形態品詞」を選択すると、富士山 固名 のように表示される[11]。

表示情報を何も選択しないと □ のような情報表示のない幅の小さなボックスを配置するので表示場所をとらずにすむ。束図の概観を得るには良い。

4.1.4 ＜サイズ変更▾ ＞

次の 3 つのサイズが変更できる。

＜辺フィールドの幅＞ 辺フィールドの幅を調節できる。束図をコンパクトに表示したい時に便利。

＜ボックスの高さ＞ 形態素ボックスとダミーボックスの高さの余白の大きさを調節できる。

＜ボックスの幅＞ 形態素ボックスとダミーボックスの幅の余白の大きさを調節できる。

なおサイズを変更すると自動的に＜描き直し＞ (4.2.7を参照) が行なわれる。

4.1.5 ＜連接コスト表示＞

束図上の辺に連接コストを表示することができる (図 16)。このボタンで表示するか否かの切替えができる。

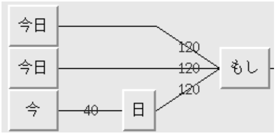


図 16: 連接コストの表示

束図が複雑な場合、数字が重なり合って見えにくくなるので、まず余計な形態素を削除 (4.2.5、4.2.6を参照) してから連接コストを表示すると良い。

4.1.6 その他の機能

＜その他の設定値▾ ＞メニューにより以下のような設定ができる。

＜品詞略称を用いる＞ 形態素ボックス上の表示情報として「形態品詞」を選び「品詞細分類」を選ばない場合、形態素ボックス上に省略形を表示することができる。

このボタンにより省略形を用るか否かの切替えができる。省略形については 5.3(p.26) を参照。

---

[10] 「形態品詞」「品詞細分類」をそのまま表示するとスペースをとるので 2 ~ 3 文字からなる省略形を用いる (例 :接尾辞. 名詞性述語接尾辞 ⇒ 名述尾)。省略形を用いない設定もできる。詳しくは 5.3(p.26) 、4.1.6を参照。

[11] 「固名」は「名詞. 固有名詞」の省略形。

< 自動で情報パネルに情報表示する > デフォルトではマウスカーソルが形態素ボックスに入っただけで情報パネル上の情報が変わるが、このボタンにより形態素ボックスをクリックした時だけ情報パネルに形態素情報を表示するように切替えられる。

< パスを盛り上げる > マウス 1 ボタンで形態素ボックスをクリックすると、最適パス・通過パス上の全形態素ボックスは色がついて盛り上がるが、このボタンにより盛り上げるか否かを切替えられる。

< コストを **10** で割る > JUMAN では単語のコスト重みに小数点以下一桁までの数値が扱えることになっているが、内部的にはコストを 10 倍して整数値として計算している。コストを内部値ではなくユーザ辞書に定義された通りに表示したいときにこのボタンを選択する。ただし、小数点以下の数値は失われる (6 『バグ』を参照)。

< 通過パスを強調 > 通過パスを強調 (色つけ&盛り上げ) するか否か切替える。

< 最適パスを強調 > 最適パスを強調 (色つけ&盛り上げ) するか否か切替える。

4.2 形態素情報とパスの選択

4.2.1 形態素の詳細情報

各形態素ボックス上にマウスカーソルが入った時[12]、情報パネル上に形態素の詳細情報が表示される。

見出し語	し	■	形態品詞	動詞	☐	最適コスト	329	☐
読み	し	☐	品詞細分類	*	☐	品詞コスト	100	☐
基本形	する	☐	活用型	サ変動詞	☐			
意味	0	☐	活用形	基本連用形	☐	本当のコスト	?	

図 17: 形態素の詳細情報 (再掲)

4.2.2 形態素情報の修正

情報パネル上のデータを変更することにより、形態素の詳細情報を修正することができる。

- 見出し語・読み・基本形・意味は修正情報を直接打ち込み (ローマ字かな変換機能により直接ひらがなを入力できる。詳しくは 4.5.1(p.20) を参照)、リターンキーまたは情報パネル上にある < 確定 > ボタンを押して修正する。
- 形態品詞・品詞細分類・活用型・活用形は情報パネル上の < 品詞と活用▼ > ボタンで選択的に変更する (図 18)。これにより文法的に矛盾の無い修正が可能である。

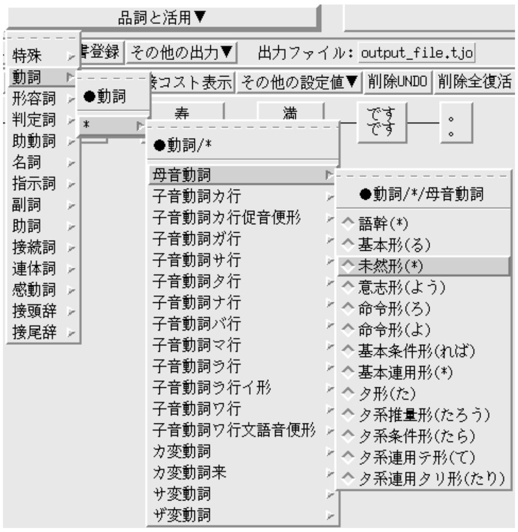
  また活用語に変更した場合、現在の活用形から自動的に基本形を推論する[13]。
- コスト関連の値は修正できない。

間違えて修正してしまったら、元の文を解析し直さない限りデータは戻らない。

---

[12] 自動情報表示モード (4.1.6を参照) でない場合は、形態素ボックスをクリックした時

[13] たとえば名詞「ミルク」を動詞の母音動詞型の語幹に変更すると、基本形は自動的に「ミルクる」となる。



出力ファイル: output_file.tjo

図 18: 品詞 ·活用情報の選択的な変更 (再掲)

4.2.3　最適パスの表示と選択

形態素ボックスをマウスの 1 (左) ボタンでクリックすると、その形態素を通る最適パス上の形態素ボックスが盛り上がって紫色になる。そしてそれらの形態素列は出力用としてシステム内部に記憶される。

また同時に、その形態素を通る全てのパス (最適パス以外) 上の形態素ボックスが空色になる。

さらに、情報パネル上に、クリックした形態素までの最適コストが、図 19のように表示される。

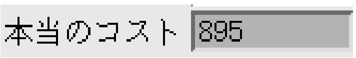

本当のコスト 895

図 19: 本当のコスト

4.2.4　＜お色直し＞

最適パスをキャンセルする。つまり何も選択されていない状態にする。紫や空色の形態素ボックスが元の色に戻るのでこの名前がついた。

マウスの 2 (中) ボタンを形態素ボックス以外の場所でクリックしても同様の効果が得られる。

4.2.5　＜削除モード＞

通常モードと削除モードを切替える。

削除モードでは、形態素ボックスをマウスの 1 (左) ボタンクリックすると、その形態素ボックスが削除され、さらにその影響で浮いてしまう形態素ボックスも削除される。図 20(a) では形態素ボックス 4 を削除すると形態素ボックス 2 が浮いてしまうので、同時に消してしまう。

4.2.6　不要な形態素ボックスを一気に消す

形態素ボックスをマウスの 2 (中) ボタンでクリックすると、マウスの 1 (左) ボタンの時と同様の効果 (4.2.3を参照) があり、さらにこの形態素に接続 (隣接ではない) できない全ての形態素

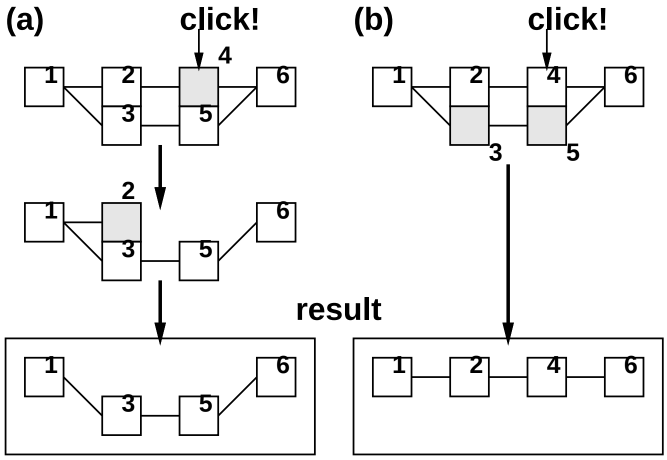


図 20: 形態素ボックスの削除

ボックスが削除される。図 20(b) では形態素ボックス 3 と 5 を通る全てのパスが形態素ボックス 4 を通らないので削除される。

実際は、これらの 2 つの方法で削除された形態素ボックスは色が黒くなるだけで、束図上に残ったままである。これらの黒くなった形態素ボックスは < 描き直し > により束図上から消すことができる (4.2.7を参照)。

4.2.7　< 描き直し >

束図を破壊してまた描き直す処理を行なう。
< 描き直し > ボタンをクリックするか、画面上の任意の場所でマウスの 3 (右) ボタンを押せばよい。

4.2.8　< 削除 UNDO > と < 削除全復活 >

削除された形態素ボックスを 1 個ずつ (< 削除 UNDO >)、また一気に全部 (< 削除全復活 >) 復活させる。 < 削除 UNDO > では一番最近に削除された形態素ボックスが復活する。

4.3　出力

4.3.1　出力ファイルの選択

< 出力ファイルの選択 > により出力先のファイルを選択できる。選択されたファイル名は出力メニュー上に表示される。
デフォルトの出力ファイル名は “output_file” となっている。

4.3.2　ファイルへの出力

ファイルへの出力に関して、 < ファイル出力 > と < 失敗マーク付き出力 >[14] の 2 つの方法がある。

---

[14] < その他の出力▾ > メニューの下にある。

<ファイル出力> 最適パス上の形態素情報を出力ファイルへ書き込む。

<失敗マーク付き出力> <ファイル出力>と同様に形態素情報を書き込むが、その前に

＊＊＊以下の文は解析に失敗しました＊＊＊

というようなマークを書き込む (上記のマークはデフォルト。マークは自由に設定できる。詳しくは5.1.3(p.23) を参照)。JUMAN の解析結果が意図したものでなかったときなどに利用する。あとで出力ファイルを見返すときに分かりやすい。

最適パスが選択されていなければファイル出力はされない (エラーメッセージがでる)。

ファイルへの出力の前に、確認のためのウィンドウが現われデータが表示される。「確認無し」の設定にしたい時は<その他の出力▼ >の下の<最適パス出力時の確認>を OFF にすればよい。

出力メニューの左端に「出力チェックマーク」というものがあり、出力を終えると変化する。新たな文を解析すれば元に戻る。

出力チェックマーク	意味 (色)
● 未出力●	まだ出力していない (赤色)
● 出力済●	すでに出力した (青色)

4.3.3 出力フォーマット

出力フォーマットは .vijumanrc で自由に変更することができる。詳しくは5.1.3を参照。

デフォルトの出力では次のようなフォーマットになっている。

見出し語/読み/基本形/品詞名/品詞細分類/活用型名/活用形名

出力例 : 出来たら できたら 出来る 動詞 * 母音動詞 タ系条件形

4.3.4 <辞書登録>

未定義語など今まで辞書に存在しなかった形態素を直接辞書に追加登録するための機能である[15]。

情報パネル上に表示されている形態素情報を JUMAN の辞書の書式でユーザ形態素辞書に書き込み、さらにコンパイルもできる。

実際の内部の処理手順について述べる。

1. まず、情報パネル上の形態素情報を JUMAN の辞書の書式にする[16]。

   (名詞 (人名 ((見出し語 リタ)(読み りた))))

2. 次にこれを引数として .vijumanrc で設定された自動辞書登録スクリプト (設定の仕方は5.1.3(p.23) を参照) を実行する。例えば設定された自動辞書登録スクリプトが/home/tatuo-y/dic_add という名前ならば、

```
/home/tatuo-y/dic_add "(名詞 (人名 ((見出し語 リタ)(読み りた))))"
```

   を実行する。

---

[15] もちろんその未定義語の品詞・活用情報は修正されているとする。

[16] このとき確認のためのウィンドウが現われデータが表示される。「確認無し」の設定にしたい時は<その他の出力▼ >の下の<辞書登録時の確認>を OFF にすればよい。

3. 自動辞書登録スクリプトの標準出力を受け取りそのまま標準出力に出力する。

以下に自動辞書登録スクリプトの例を示す。

```
#!/bin/csh -f
cd /home/tatuo-y/juman/dic
echo $1 >> Hozon
echo $1 > my.dic
makeint my.dic >& err_msg_1
makehd my.int  >& err_msg_2
cat err_msg_1 err_msg_2
```

このスクリプトにより、ディレクトリ /home/tatuo-y/juman/dic に JUMANTREE.dbm.pag 等のコンパイルされた辞書ができる。既にある場合は追加登録される。

JUMAN は .jumanrc により、参照する複数の辞書を指定できるので、この辞書も指定しておく。以下に .jumanrc 内での辞書の設定例を示す (/home/nltools/juman2.0/dic/JUMANTREE は標準の辞書である)。

```
(辞書ファイル
        /home/nltools/juman2.0/dic/JUMANTREE
        /home/tatuo-y/juman/dic/JUMANTREE )
```

これにより、直後の解析[17]で今登録した単語が利用できる。

`echo $1 >> Hozon` により、ファイル Hozon (今まで追加登録したユーザ辞書項目が蓄積されているファイル) に新たに登録した辞書項目を追加する。

`makeint`, `makehd` はその処理過程をエラー出力するが、これは Tcl/Tk にとっては非常に困る (本当のエラーと認識し、エラー処理が始まる)。そこで、エラー出力を一度ファイルに落して (`>& err-msg...`)、それを標準出力に出すようにしている (`cat err_msg...`)。これにより、処理 3 でコンパイル状況が分かる。

辞書のコンパイルについては文献 [1] を参照されたい。

4.3.5 <単語出力>

<辞書登録>と同様に、未定義語として解析された文字列に正しい形態素情報を指定して、記録しておくための機能である。

現在、情報パネルに表示されている形態素情報をパスの出力フォーマットでファイル出力する。出力先のファイルは、.vijumanrc で設定できる。詳しくは 5.1.3(p.23) を参照。デフォルトでは “new_words” となっている。

4.3.6 <束図の出力>

<その他の出力▾ >メニューの下の<束図の出力>を選ぶと現在ディスプレイに表示されている束図が画像データとして出力される。

より正確に言うと、キャンバスのウィンドウ ID を引数として vijuman_out というシェルスクリプトが実行される。

---

[17] ただし解析前に< JUMAN 再起動>ボタンを押す必要がある。

デフォルト (配布時) の vijuman_out では、まず束図の画像データを PostScript に変換して ghostview で表示するようになっている。このマニュアル中の束図はこの機能で取り込んだものである。

中身は次のようになってる。

```
#!/bin/csh -f
xwd -id $1 -out /tmp/PIC$1.xwd
convert /tmp/PIC$1.xwd /tmp/PIC$1.ps
ghostview /tmp/PIC$1.ps
rm /tmp/PIC$1.ps /tmp/PIC$1.xwd
```

自分専用の vijuman_out を作って利用することもできる。その場合パスの指定が必要となるが、詳しくは 5.1.3(p.23) を参照されたい。

4.4 入力と解析

4.4.1 文の直接入力

入力スペースに直接文を入力すれば良い。ローマ字かな変換機能により直接ひらがなを入力できる (詳しくは 4.5.1(p.20) を参照)。またはマウスのカット&ペースト機能で入力する。

＜クリア＞により入力スペースを空にできる。

4.4.2 ファイルからの入力

＜入力ファイルの選択＞により入力ファイルを選択する。そして＜１文読み込み＞によりそのファイルの次の一行が、また＜１文戻り＞により前の一行が読み込める。

読み込まれた行は入力スペースに表示される。また入力パネルのメニューの右側に入力ファイル名と今読み込んだ行番号が表示される (図 21)。行番号が表示されている所に直接数字を入力することにより、そのファイル内の任意の行を読み込むことが出来る。

図 21: ファイル情報

4.4.3 ＜メモ▾ ＞

＜ Save ＞と＜ Load ＞の二つの機能がある。＜ Save ＞により現在の入力ファイル名と行番号を保存することができる。保存したデータは＜ Load ＞により読み込むことができ、保存した状態から作業が続けられる。データは ~/.memo.vijuman に保存されている。

4.4.4 ＜解析する＞

このボタンにより、入力スペース上の文を JUMAN で解析して束図として表示する。入力スペース上でリターンキーを押しても同様。

ただし JUMAN で解析するには JUMAN サーバ (`juman_pl`) が起動されている必要がある。詳しくは 4.5.2(p.20) を参照。

4.4.5 ＜部分再解析＞

文を解析した後、その1部分だけ再び解析したいことがある。

入力スペース上の文から、マウスで再解析したい部分だけセレクト (図22) して＜部分再解析＞ボタンを押せば良い。その部分だけJUMANで解析される。そのときのコスト幅は自由に設定できる (4.4.7を参照)。


これは美寿満です。

図22: マウスによるセレクト (再掲)

4.4.6 ＜部分無解析＞

＜部分再解析＞とほぼ同じだが、解析は行なわれずセレクトした部分を未定義の1単語とする。これにより任意に単語の合体・分割ができる。

例として『こんにちは。』の解析結果を用いて説明する。

こんにちは ─ 。

このとき『にち』をセレクトすると、

こん ─ にち ─ は ─ 。

というように、『こんにちは』が3つに分割される。またこの段階で『ちは。』をセレクトすると、

こん ─ に ─ ちは。

というように、分割と合体が同時に行なわれる。このような分割・合体は＜部分再解析＞でも行なわれている。

部分再解析と部分無解析の注意点を以下に上げておく。

解析後の最適コストの値が実際と違う 形態素ボックス上にコストを表示すると分かる。

解析後の連接コストの値が0になる

解析する範囲を選ぶ際に元の文によってバグが生じることがある 6に詳しく述べてある。

分割したとき、読みが冗長になることがある 6に詳しく述べてある。

4.4.7 コスト幅の変更

図23のように、情報パネル上でJUMANで解析する際のコスト幅を設定できる[18]。値はメニューで選んでも直接打ち込んでも良い。ただし、一般解析用のコスト幅の変更後には＜**JUMAN** 再起動＞ボタンを押して **JUMAN** を立ち上げ直す必要がある。デフォルトのコスト幅はJUMANのリソースファイル (.jumanrc) で設定されている値を用いる。

---

[18] 実際は、JUMANを実行する直前に .jumanrc のコスト幅を書き変えている。

また、部分再解析専用のコスト幅も設定・変更できる。デフォルトは 300 だが、.vijumanrc 中で初期値を設定できる。詳しくは 5.1.4(p.25) を参照。部分再解析用コスト幅の変更後には< **JUMAN** 再起動>の必要はない。

通常の解析ではコスト幅を低めにしておき、望みの解析結果が得られない部分に対してのみ高めのコスト幅 (部分再解析専用) で部分再解析を行なうのが一般的な手法である。

コスト幅

一般解析用 100

部分解析用 300

図 23: コスト幅の変更

4.5 その他

4.5.1 ローマ字かな変換

入力スペースと情報パネル上の見出し語・読み・基本形のスペースではローマ字かな変換機能により直接ひらがなを入力できる。

例えば、入力スペース上で『kaNpokkuri』と打ち込むと、順次変換されて『かんぽっくri』と表示される[19]。ここでシフトキーやコントロールキーを押せば『かんぽっくり』となって変換が完了する。

変換テーブルは訓令式・ヘボン式を意識せず思いついたもので作ったので意図した「かな」に変換されない事もあるかも知れない。

`vijuman -rk` を実行すると、ローマ字かな変換規則の一覧が出力される。

アルゴリズムの都合上、入力誤り (かなに変換できないローマ字の入力) があるとその先の入力は変換されない。

4.5.2 JUMAN サーバの起動

JUMAN で解析するには JUMAN サーバ (`juman_pl`) が起動されている必要がある。`juman_pl` が起動されていないにもかかわらず解析を行なおうとするとエラーとなる。

このような場合の`juman_pl` の起動の仕方について説明する。

まずシステム内でデフォルトとして設定されているホストマシン名を調べる必要がある。

`vijuman -ph` を実行するとシステムが使用するデフォルトのポート番号[20]とホストマシン名が表示される。

```
cosmos % vijuman -ph
Port = 16010, Host = violet
```

---

[19] 最後の『i』がそのままなのは、「本当は『り』に変換されているのだが表示が追い付いていない」と考えて頂きたい。

[20] ポート番号は滅多に変更されないと思われるので、これ以降では無視する。

次にそのマシンで、`juman_pl` を起動する。

```
cosmos % rsh violet juman_pl &
```

これで美寿満での解析が可能となる。

またデフォルトのマシン以外で`juman_pl` が起動されていることが分かっていれば、コマンドラインオプションで指定して利用することができる。

例えば、デフォルトのホストマシン名は`violet` だが、`fennel` というマシンで`juman_pl` が立ち上がっているということが分かっている場合は次のようにして美寿満を起動すれば良い。

```
cosmos % vijuman -H fennel
```

4.5.3 コマンドラインオプション

vijuman では以下のコマンドラインオプションが利用できる。

-i *filename* :入力ファイルの指定

-r *filename* :美寿満のリソースファイルの指定

美寿満のリソースファイル (デフォルトは ~/.vijumanrc) を指定できる。

-j *filename* :JUMAN のリソースファイルの指定

JUMAN の立ち上げに用いるリソースファイル (デフォルトは ~/.jumanrc) を指定できる。.vijumanrc でも設定値 jrc_file で指定できる。5.1.3(p.23) を参照。ただし、コマンドラインオプションより .vijumanrc での指定の方が優先される。

-l *language* :メニュー表示に用いる言語の指定

『日本語 (jp)』か『英語 (en)』を指定できる。デフォルトでは日本語表示。.vijumanrc でも設定値 lang で指定できる。5.1.4(p.25) を参照。ただし、コマンドラインオプションより .vijumanrc での指定の方が優先される。

-ph :美寿満で利用する JUMAN サーバ (`juman_pl`) のポート番号とホストマシン名の設定を表示。4.5.2(p.20) を参照。

-H *hostname* :`juman_pl` のホストマシンの指定。4.5.2(p.20) を参照。

-P *port_number* :`juman_pl` のポート番号の指定。

-rk :ローマ字かな変換規則の一覧を出力する。4.5.1(p.20) を参照。

-m :メモしたところから始める。

~/.memo.vijuman に記録されている入力ファイル名と行番号を起動時に Load する。メモ機能については 4.4.3(p.18) を参照。

-t :起動時にタイトルを表示するかどうかの切替え。

-ht :品詞略称テーブルを表示する。

5　カスタマイズ

5.1　リソースファイル “.vijumanrc”

カスタマイズのための .vijumanrc というファイルがあり、自分のホームディレクトリの下に置くことにより利用できる[21] 。

書式は次の通り。

```
set [設定値名] [設定値]
```

必ず半角スペース 1 文字以上で区切る。また、行頭に # があるとその行はコメントと見なされ無視される。

5.1.1　束図表示関連の設定値

- 表示形態を指定する。(0-コスト順, 1-中寄せ, 2-すっきり配置, 3-すっきり中寄せ)

  ```
  set hyouji_type 2
  ```

- 形態素ボックス上に表示させる情報を指定する。内部コード化されているので直接書き換えるより、実行中に表示情報を指定 (4.1.3(p.11)) して、＜設定値保存＞ (p.26) を用いるのが無難。

  ```
  set list_of_info_on_box 1000100000
  ```

- 省略形を用いるか否か指定する。(0-no, 1-yes)

  詳しくは 4.1.3(p.11), 5.3を参照。

  ```
  set ryakusyou_mode 1
  ```

- 自動で情報パネルに情報表示するか否か指定する。(0-no, 1-yes)

  詳しくは 4.1.6(p.12) を参照。

  ```
  set auto_show_info 1
  ```

- 立ち上げ時の辺フィールドの幅をピクセル単位で指定する。

  ```
  set edge_field_width 60
  ```

- 形態素ボックスの縦方向の余白をピクセル単位で指定する。

  ```
  set box_height_plus 16
  ```

- 形態素ボックスの横方向の余白をピクセル単位で指定する。

  ```
  set box_wide_plus 16
  ```

---

[21] .vijumanrc がホームディレクトリに無い場合は、美寿満の起動時にデフォルトの .vijumanrc がコピーされる。

5.1.2 パス関連の設定値

- クリックした形態素ボックスを通る最適パス上の形態素ボックスの色を指定できる。spdabg はマウスカーソルがボタンに入った時の背景色 (activebackground)。

  ```
  set spdbg thistle2
  set spdabg thistle3
  ```

  強調 (色つけ&盛り上げ) するか否かも指定できる。(0-no, 1-yes)

  ```
  set show_best_path_flag 1
  ```

- 同様に通過パス[22]上の形態素ボックスの色も指定できる。

  ```
  set apdbg skyblue2
  set apdabg skyblue3
  ```

  強調するか否かも指定できる。(0-no, 1-yes)

  ```
  set show_all_path_flag 1
  ```

- 最適パス・通過パス上の形態素ボックスを盛り上げるか否か指定する。(0-no, 1-yes)

  ```
  set emboss_path 0
  ```

- 立ち上げ時に削除モードか否かを指定する。(0-通常モード, 1-削除モード)

  ```
  set delete_mode 0
  ```

- 立ち上げ時に連接コストを束図上に表示するか否かを指定する。(0-表示しない, 1-する)

  ```
  set show_edge_cost 0
  ```

5.1.3 出力関連の設定値

- 出力フォーマット・出力の区切り記号・出力先ファイルの指定。

  ```
  set output_format 見出し語/読み/基本形/品詞名/品詞細分類/活用型名/活用形名
  set output_delimiter ,
  set output_file output_file
  ```

  出力可能なデータ (全１３個) を以下に示す。

---

[22] クリックした形態素ボックスを通る全てのパス (最適パス以外)

データの分類	出力可能なデータ			
内部処理用	識別子	前接形態素の識別子のリスト		
表層と意味	見出し語	読み	基本形	意味情報
品詞と活用	品詞名	品詞細分類	活用型名	活用形名
コスト関連	最適コスト	品詞コスト	連接コスト	

例えば見出し語·読み·品詞名のみを出力したければ、output_format を見出し語/読み/品詞名 と設定すれば良い。必ず / で区切ること。デフォルトの出力フォーマットは、

見出し語/読み/基本形/品詞名/品詞細分類/活用型名/活用形名

である。

output_delimiter により出力時にデータを区切る記号を指定できる。上記の設定 (,) では出力は次のようになる。

今日, こんにち, 今日, 名詞, 時相名詞,*,*

区切り記号に半角スペース·タブ·改行を指定したいときは Space, Tab, Return という文字列で設定する。

例： `set output_delimiter Space`

output_file により出力先のファイルが指定できる。

- 失敗出力用のメッセージを指定する。

```
set fail_output_message "---以下の解析結果はちょっと怪しい---"
```

- 新単語出力用のファイルを指定する。

```
set new_word_file new_words
```

- 自動辞書登録スクリプトを指定する。自動辞書登録スクリプトについては 4.3.4(p.16) を参照。

```
set dic_append_script /home/tatuo-y/bin/dic_proc.perl
```

- デフォルトのでなく、自分で作った束図出力スクリプト (ファイル名は必ず vijuman_out でなければならない) を使いたい時は、そのファイルを置いたディレクトリを指定する。束図出力スクリプトについては 4.3.6(p.17) を参照。

```
set vijuman_out_path /home/tatuo-y/tk
```

- 最適パスをファイルに出力するとき確認するか否か指定する。(0-no, 1-yes)

```
set path_output_preview 1
```

- 単語を辞書登録するとき確認するか否か指定する。(0-no, 1-yes)

```
set dic_output_preview 1
```

5.1.4 その他の設定値

- 全体の背景色を指定する。dbg は普通の背景色 (background)、dabg はマウスカーソルがボタンなどに入った時の背景色 (activebackground)。

```
set dbg pink
set dabg yellow
```

- JUMAN の立ち上げに用いるリソースファイルを指定する。デフォルトは ~/.jumanrc

```
set jrc_file ~tatuo-y/juman_test/.jumanrc.org
```

- メニュー表示に用いる言語を指定する。(jp または en 。デフォルトは jp)

```
set lang en
```

- 入力パネルを別ウィンドウにするかどうか指定する。(0-no, 1-yes)

```
set separate_input_panel 0
```

- 部分再解析用のコスト幅を指定する。

```
set cost_width2 300
```

- デフォルトのでなく、自分で作った「品詞名略称テーブル」(5.3を参照) を使いたい時は、そのファイルを指定すればよい。

```
set hinsi_table_path /home/tatuo-y/tk/hinsi_table
```

- 漢字 1 文字の大きさを指定する。フォントの指定ではない。

```
set char_size 16
```

- 全てのコストを 10 で割るか否か指定する。(0-no, 1-yes)

```
set cost_div_10 1
```

- 起動時にタイトルを表示するか否か指定する。表示をやめてもあまりスピードは変わらない。(0-no, 1-yes)

```
set title_window 1
```

5.2 ＜設定値保存＞

表示形態や形態素ボックスの幅などをシステム実行中に変更して、次からも同じ設定で実行したいとき、わざわざ .vijumanrc を書き換えるのは面倒である。そこでこの＜設定値保存＞を実行することにより、以下に示す .vijumanrc 中の設定値を現在の設定値に合わせて自動的に書き換えてくれる。

```
edge_field_width      hyouji_type           list_of_info_on_box
delete_mode           output_file           box_height_plus
box_wide_plus         emboss_path           cost_div_10
show_edge_cost        path_output_preview   dic_output_preview
show_all_path_flag    show_best_path_flag   ryakusyou_mode
auto_show_info        cost_width2
```

5.3 品詞名略称テーブル

ここでは品詞名略称テーブル (hinsi_table) について説明する。

長い品詞名があると形態素ボックスの幅をとってしまい、結果的に束図の幅がとてつもなく大きくなってしまうことがある。そこで形態素ボックス上に品詞名を表示する時には短い略称を用いることができる (4.1.3(p.11) を参照)。hinsi_table はそのための変換規則を記述するファイルである。

略称テーブルのフォーマットは次の通り。

[元の品詞名[23]][1 個以上の半角スペース][略称]

元の品詞名と略称の間には半角スペース以外の文字を入れてはいけない。略称を元の品詞名と同じにしたい場合は何も記述しなくてよい

以下にデフォルトの品詞名略称テーブル (hinsi_table) を示す (`vijuman -ht` でも見ることができる)。“動詞” や “句点” などは品詞名が短く、略称を必要としないので記述されていない。

普通名詞	普名	サ変名詞	サ名
固有名詞	固名	形式名詞	形式名
副詞的名詞	副名	時相名詞	時名
名詞形態指示詞	名指示	連体詞形態指示詞	連指示
副詞形態指示詞	副指示	様態副詞	様副
程度副詞	程副	量副詞	量副
頻度副詞	頻副	時制相副詞	時副
陳述副詞	陳副	評価副詞	評副
発言副詞	発副	格助詞	格助
副助詞	副助	引用助詞	引助
名詞接続助詞	名接助	述語接続助詞	述接助
終助詞	終助	名詞接頭辞	名接頭
動詞接頭辞	動接頭	イ形容詞接頭辞	イ接頭
ナ形容詞接頭辞	ナ接頭	名詞性述語接尾辞	名述尾
名詞性名詞接尾辞	名名尾	名詞性名詞助数辞	名助辞
形容詞性述語接尾辞	形述尾	形容詞性名詞接尾辞	形名尾
動詞性接尾辞	動接尾	未定義語	未定義

自分専用の hinsi_table を作って利用することもできる。詳しくは 5.1.4を参照。

---

[23] 品詞細分類がある場合は品詞細分類名

6　バグ

部分再 (無) 解析周辺のバグ

1. マウスで再 (無) 解析したい部分を選ぶ時、もし元の文がスペースで始まっていたり半角文字が混じっていたりするとうまく再 (無) 解析できないことがある。原因は Tcl/Tk と C(JUMAN) での全角文字の数え方が違うことによる (Tcl/Tk では全角半角問わず全て 1 文字として数えるが、JUMAN では全角文字を 2 文字、半角文字を 1 文字と数える)。
2. 全角スペース混じりの文で部分再 (無) 解析するとエラーが出ることがある。これは JUMAN が解析時に全角スペースを無視することによる。
3. 再 (無) 解析後の最適コストの値は解析前と整合性がない。また、連接コストの値は 0 になる。 バグというより手抜きである。
4. 再 (無) 解析後に、浮いてしまう形態素 (4.2.5を参照) が残ることがある。
5. 単語を分割するとき「読み」に問題がある。例えば「解析」(かいせき) を「解」と「析」に分割するとき、「か-いせき」「かい-せき」「かいせ-き」の 3 パターンが考えられ、読みをどのように割り振るか非常に困る。そこで「解」「析」ともに読みは「かいせき」としている。

   「解」や「析」をいちいち JUMAN の辞書にかけて読みを調べても良いのだが、「多い分には、マウスで消せばいい」という考えに基づき今のところ読みは冗長になっている。 バグというより手抜きである。

コストを 10 で割ることによる弊害 本システムでは、JUMAN の解析結果を読み込む時に全ての内部コスト (最適・連接・品詞) を 10 で割っている (1 未満切捨て) ので、例えば内部コスト 165 の場合は 16.5 にはならず、16 になってしまい、文末の形態素ボックスをクリックした時に計算される「本当のコスト」の値が、そこの「最適コスト」と少し違って来る場合がある。

— 例文 —

- 1 いじょう 2 みまん
- じょじょに休めば、ますます疲れる。
- かならずしも、なんでも出来る訳ではない。

たとえば「1 いじょう 2 みまん」をコスト幅 0 で解析した結果をみてみると、「みまん」の最適コストは 395 だが 本当のコストは 394 となる。これは「いじょう」「みまん」の品詞コストがもともと 165 であったのが、10 で割って 16 (1 未満切捨て) になったのが原因である。

いまのところ、どうしようもないので、コストを 10 で割るか否かを選べるようにしてある (< その他の設定値▾ > の下の < コストを 10 で割る > ボタン)。

すっきり中寄せ 真中に寄らず、下方向にずれて配置されることがあり、キャンバスをはみ出すこともある。

参考文献

[1] 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, July 1994

- URL: http://www.aist-nara.ac.jp/IS/TechReport2/report/94025.ps で入手できる。

[2] Larry Wall and Randal L. Schwartz(著) 近藤嘉雪 (訳), Perl プログラミング, ソフトバンク株式会社, 1993

[3] 宮田重明・芳賀敏彦, Tcl/Tk プログラミング入門, オーム社, 1995

[4] John K. Ousterhout, Tcl and the Tk Toolkit, Addison-Wesley Publishing Company, 1994

[5] John K. Ousterhout(著) 西山芳幸・石曽根信 (訳), Tcl & Tk ツールキット, ソフトバンク株式会社, 1995

- 文献 [4] の邦訳

[6] B. Schneiderman(著) 東基衛・井関治 (監訳), ユーザ・インタフェースの設計 やさしい対話型システムへの指針 第2版, 日経BP出版センター, 1995

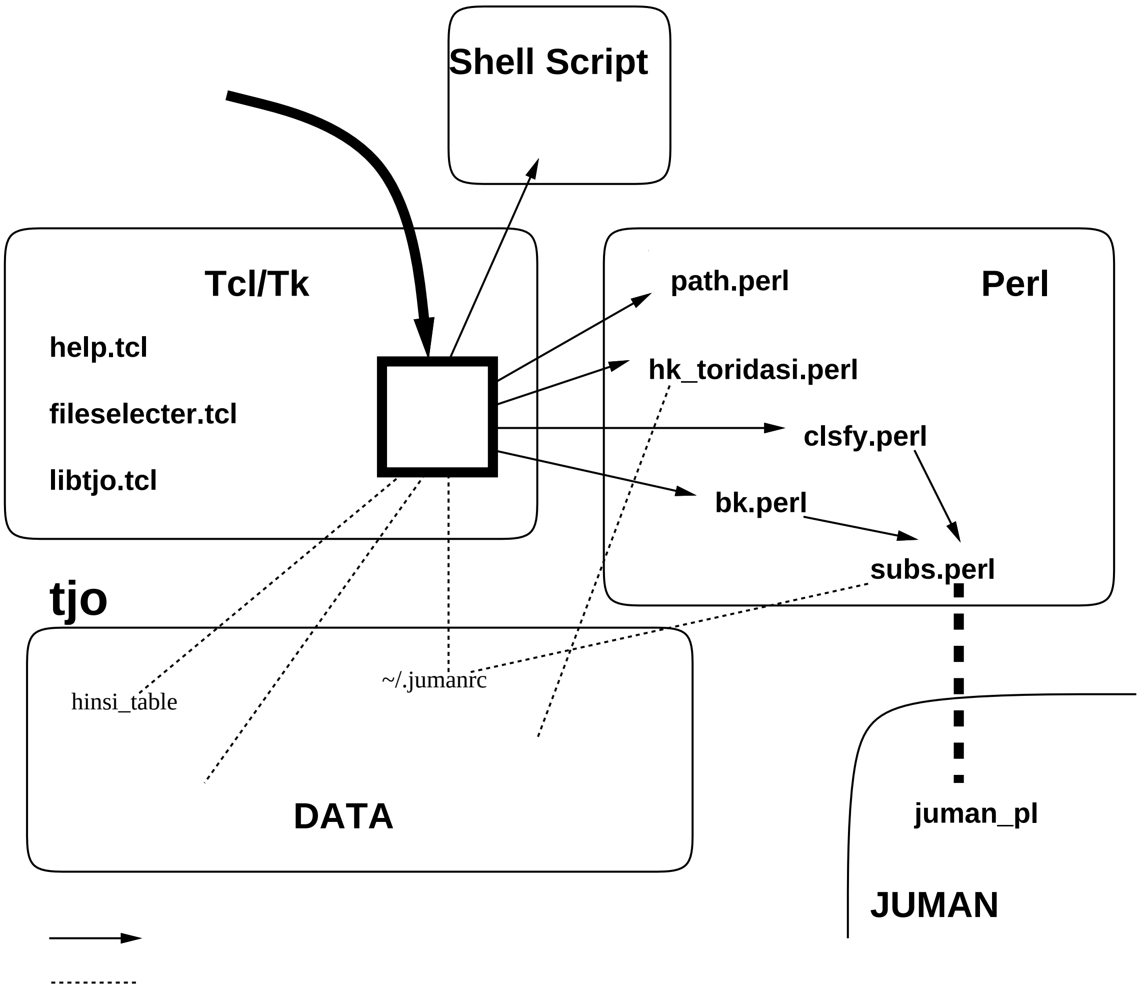
[7] 益岡隆志・田窪行則, 基礎日本語文法 —改定版—, くろしお出版, 1992

付録

A　Version 0.9 からの主な変更点

- コマンドラインオプションを新設。
- <表示形態▾ >に<すっきり中寄せ>を追加。
- 解析毎に立ち上げ直していたJUMANを連続解析可能にした。コスト変更時用に< JUMAN再解析>を新設。
- メモ機能を新設。<メモ▾ >ボタンを新設。
- 今までは、品詞細分類が定義されていない場合、品詞細分類は品詞名と同じになっていたが、C版JUMANにあわせて * にした。
- 基本形の推論機能の実装。品詞と活用を変更した時に「基本形」を見出し語と活用形から自動的に推論する。
- 連接コスト表示機能を実装。連接コストが辺(edge)上に表示されるようになった。それにともない情報パネル上から連接コスト表示欄を削除。また表示切替え用に<連接コスト表示>ボタンを新設。
- 最適パスデータ・辞書登録データの出力を事前に確認するウィンドウを新設。
- 形態素情報の修正を確定するための<確定>ボタンを新設。
- 入力ファイルから一行まえのデータを読み込む< 1文戻り>ボタンを新設。
- メニューの英語表記モード(`vijuman -l en` で起動)を新設。
- 文入力・情報修正用にローマ字かな変換機能を実装した。
- 形態素ボックス上の情報表示を一段から複数段にした。これにともない、形態素ボックス上の情報を縦書きで表示する<縦書きモード>を廃止した。
- さらに<表示情報▾ >を廃止。ボックス上に表示する形態素情報は情報パネルで選ぶようにした。
- はずかしいバグをこっそり修正した。

B　システム構成



vijuman	メイン (Tcl/Tk で書かれている)
help.tcl	Vi JUMAN のウィンドウヘルプ
fileselecter.tcl	ファイル選択モジュール
libtjo.tcl	色々な処理を行なう．
path.perl	最適パスの選択を行なう．
hk_toridasi.perl	JUMAN.{grammar,katuyou,kankei} から品詞情報と活用情報を体系的に取り出す．
clsfy.perl	JUMAN による解析結果をクラス分けする．
bk.perl	部分再解析・部分無解析を行なう．
subs.perl	juman_pl との通信を行なう．
vijuman_out	束図を PostScript 形式で出力するシェルスクリプト
hinsi_table	品詞略称テーブルファイル
~/.vijumanrc	Vi JUMAN のオプション設定ファイル