

Egg V4 (Tamago Next Generation)

片山 善夫
株式会社PFU 第三研究室
kate@pfu.co.jp

1999年12月18日

1 はじめに

ご存知のように、日本語エディター NEmacs の日本語入力メソッドとして開発された Egg は、多言語エディター Mule に引き継がれ、日／中／韓国語入力メソッドへ発展しました。Mule は GNU Emacs を多言語化して作られたのですが、その多言語機能が GNU Emacs 20 に採り入れられ、GNU Emacs 自身が多言語エディターとなりました。しかし、一部の多言語機能は未だインプリメントされていないものがあり、Egg もその一つに含まれています。このため、電総研の Mule プロジェクトの一部として、電総研以外の組織からの協力者も参加して、GNU Emacs 20 用 Egg の開発が行なわれています。今回の開発は完全に新規に行なわれ、Egg Version 4 として生まれ変わりました。

現在はβテストの準備中で、たまご(たくさんまたせて ごめんなさい)状態となっています。今後、仕様等が大きく変わる可能性があることを御了承下さい。

2 主な特徴

Egg V4 の主な特徴には、

- 完全 elisp 化
- 多言語機能の強化

があります。

2.1 完全 elisp 化

Egg は、かな漢字変換サーバーのフロントエンドとして動作します。従来は変換サーバーとの通信に、かな漢字変換システム (Wnn、SJ3) のライブラリーを用いていました。このため、

- ライブラリーをリンクするために再コンパイルが必要
- ライブラリー中で SIGSEGV 等のシグナルが発生すると Emacs がクラッシュする
- Wnn と SJ3 のいずれか一方しか使えない

などの問題点がありました。三番目は「日本語入力には SJ3 を使い、中国語入力には Wnn を使う」といったことができないことです。

Egg V4 では、ライブラリーを用いていた部分も含めて、すべてを elisp で記述することにより、上記の問題点を解消しました。また、日本語入力だけしか行なわない場合でも、Wnn と SJ3 を随時切替えて使うことも可能です。

2.2 多言語機能の強化

元々 Egg は日本語入力メソッドとして開発されたため、中／韓国語入力メソッドとしては機能不足でした。特に問題となっていた点には、

- 中国語繁体字が入力できない
- 中／韓国語入力への切替が面倒(メニューモードしかない)
- 中／韓国語の“全角”英数字が入力できない

などがあります。Egg V4 では、terver を使った中国語繁体字入力をサポートし、表 1 の入力モードを用意しました。入力モードの切替えは C-x C-m key1 で行ないます。フェンス中では ESC key2 でも切替えられ

モード名	key1	key2	入力モード	モード名	key1	key2	入力モード
言語共通							
down	q	q	ASCII	up	Q	Q	ASCII (大文字)
日本語							
hira	h	C-h	平仮名	kata	k	C-k	片仮名
han-kata	x	C-x	半角片仮名				
zenkaku-down	z	z	全角英数字	zenkaku-up	Z	Z	全角英数字 (大文字)
簡体字中国語							
pinyin-cn	C-p	C-p	拼音 (全拼)	erpin-cn	C-e	C-e	二拼 (双拼)
zhuyin-cn	C-z	C-z	注音				
qianma	C-q	C-q	銭碼	wubi	C-w	C-w	五筆
quanjiao-down-cn	C-d	C-d	全角英数字	quanjiao-up-cn	C-u	C-u	全角英数字 (大文字)
繁体字中国語							
pinyin-tw	P	P	拼音 (全拼)	erpin-tw	E	E	二拼 (双拼)
zhuyin-tw	C	C	注音				
quanjiao-down-tw	D	D	全角英数字	quanjiao-up-tw	U	U	全角英数字 (大文字)
韓国語							
hangul	H	H	ハングル				
jeonkak-down	j	j	全角英数字	jeonkak-up	J	J	全角英数字 (大文字)

表 1: 入力モード一覧

ます。

Egg V4 は、更に多言語入力メソッドとしての機能も強化しています。多言語テキストには、複数の言語で併記したもの(図 1)と、一文中に複数の言語が混在したもの(図 2)があります。

テクニカルコンファレンスへようこそ。 欢迎参加技术会议。

図 1: 多言語テキストの例(その1)

韓国語で「こんにちは」は「안녕하세요」です。

図 2: 多言語テキストの例(その2)

図 2 のタイプの多言語テキストの入力で頻繁に入力言語の切替が起こるために、入力言語の切替が簡単な多言語入力メソッドが求められます。日本語や中国語のように漢字を使う言語ではかな漢字変換操作が必要

となりますが、従来の入力メソッドでは各言語毎にかな漢字変換を行なう必要がありました。このため、図2のタイプの多言語テキストの入力は非常に複雑でした。また、このようなテキストでは、ベースとなる言語に他の言語が挿入された形になりますから、ベース言語へ簡単に戻れることも必要です。Egg V4 では、異なる言語の読みを入力しておいて、まとめてかな漢字変換を行なえるようにしています。図3は、Egg V4 による日中混在テキストの漢字変換の例です。

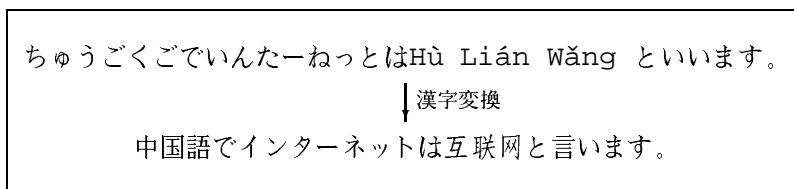


図 3: 多言語テキストの漢字変換例

更に、1キーで入力言語を切替え、C-g で復帰するようにカスタマイズすることも可能となっています。ひらがな入力時に、q で ASCII 入力へ切替え、C-g でひらがな入力に戻るのと同様です。私の環境では、

- C 拼音入力(簡体字中国語)
- F 拼音入力(繁体字中国語)
- H ハングル入力(韓国語)

としています。

3 Egg V4 の使い方

基本的な使い方は従来とあまり変わりません。ここでは、従来と異なる点について述べます。

3.1 Egg の起動

Mule では入力メソッドの制御 (on/off、モードラインの変更、etc.) は、各々の入力メソッドが行なっていましたが、Emacs 20 では leim が入力メソッドの制御を行ないます。Egg V4 も、この leim の制御下で動作します。leim による入力メソッドの on/off は C-\ で行ないます。デフォルトの入力メソッドが定義されていない時は、入力メソッド名を聞いてきます。入力メソッドを変更する場合は C-u C-\ で、この時も入力メソッド名を聞いてきます。表 2 は、Egg V4 で用意している入力メソッド名です。

日本語			
japanese-egg-wnn	平仮名入力 (Wnn)	japanese-egg-sj3	平仮名入力 (SJ3)
簡体字中国語			
chinese-gb-egg-wnn-py	拼音入力 (Wnn)	chinese-gb-egg-wnn-zy	注音入力 (Wnn)
chinese-gb-egg-wnn-qm	銭碼入力 (Wnn)	chinese-gb-egg-wnn-wb	五筆入力 (Wnn)
繁体字中国語			
chinese-cns-egg-wnn-py	拼音入力 (Wnn)	chinese-cns-egg-wnn-zy	注音入力 (Wnn)
韓国語			
korean-egg-wnn	ハングル入力 (Wnn)		

表 2: 入力メソッド名一覧

入力言語を切替えるのに、入力メソッド自体を切替える方法と入力モードを切替える方法があることに気づかれたと思います。両者の差は、出力されるメッセージの言語も変わるかわからないかだけしかありません。

3.2 .eggrc

Egg V4 の .eggrc は、従来のそれと異なっています。これは、

- Wnn と SJ3 を統合したこと
- 関数名が egg、wnn、sj3 などのプリフィックスで始まるようにした

ことによります。従来の .eggrc をそのまま使うようにカスタマイズすることもできます。

3.3 カスタマイズ

Egg V4 では、Emacs 20 の customize-group によってカスタマイズできるようにしています。M-x customize-group RET egg RET とタイプするとカスタマイズ画面(図 4)になりますので、メニューにしたがってカスタマイズを行なって下さい。

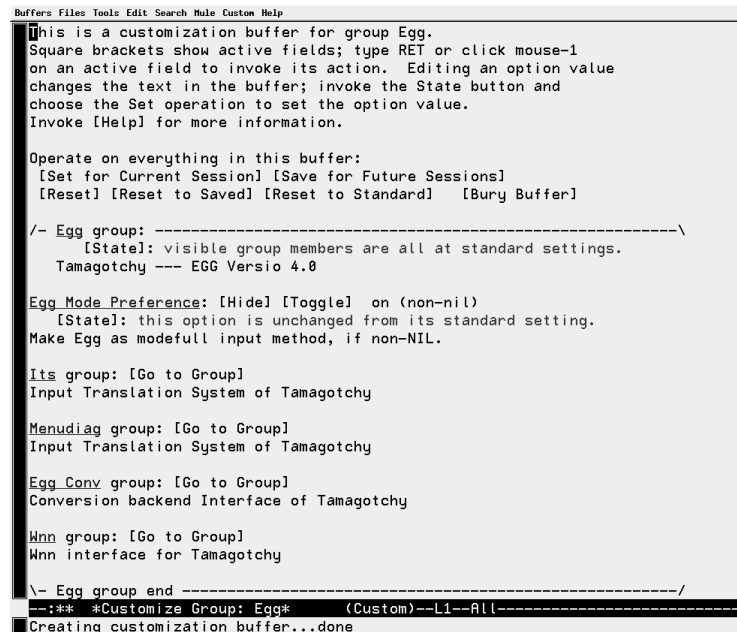


図 4: Egg のカスタマイズ初期画面

4 主な機能強化

4.1 フェンスがいっぱい!?

Egg V4 では、フェンスモード(エディットモード)や変換モードの途中でフェンスの外へ出ることが可能になっています。その状態でタイプすると、新たなフェンスが作られます(図 5)。入力中に他の部分を直したり、他のバッファを見たりすることができます。また、フェンスモードで yank (C-y) ができるようにしていますので、cut and paste してくることもできます。

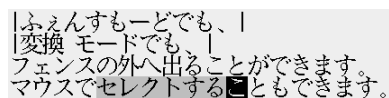


図 5: マルチフェンス

4.2 メニューモード一覧表示

メニューモードに全候補の一覧表示機能を追加しました。従来は、ミニバッファに1行ずつ表示する機能しかありませんでしたので、単漢字変換のように候補数が多い時は候補を探すのが大変でした。メニューモードで?をタイプすることによって、一覧表示ウィンドウが表示されます(図6)。このウィンドウからの選択はマウスによって行ないます。

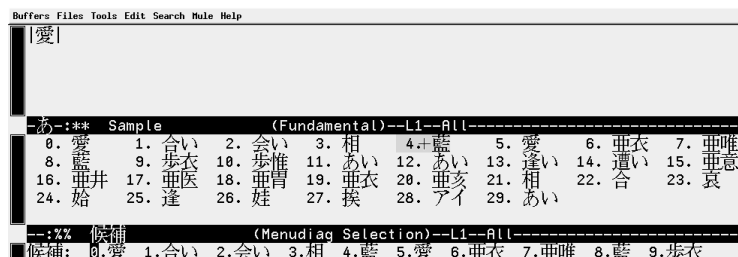


図6: 一覧表示ウィンドウ

かな漢字変換の候補選択でメニューモードに入るには、従来通りM-s (大文節変換)、M-z (小文節変換) をタイプします。この時、代わりにM-C-s、M-C-zをタイプすると、一覧表示ウィンドウも表示されます。

4.3 再変換

変換サーバーにWnnを使っている場合、変換モード中にC-rをタイプすると、カーソルがある文節が再変換されます。これは、表示されている文字列を「読み」として、再度変換を行なう機能です。部首検索辞書による漢字入力(図7)では、「部首の読み」→「部首漢字」→「漢字」の二段階の変換を行ないますが、この二段目の変換に再変換機能を使います。

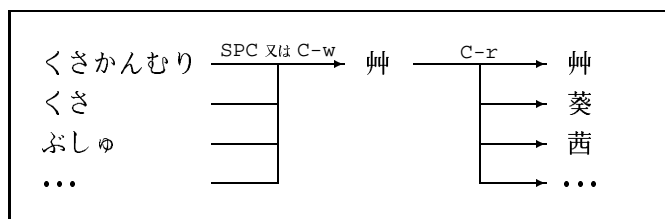


図7: 部首検索辞書による漢字入力

4.4 逆変換

変換サーバーにWnnを使っている場合、(再)変換コマンドにC-uプリフィックスを付けると逆変換(「漢字」から「読み」へ変換)になります。漢字の読みを調べたい時は、フェンス中へ漢字をyankしてきて、C-u SPC又はC-u C-wとタイプします。あるいは、読みへ変換したい部分をregion指定しておいて、C-u M-x egg-convert-region RETとタイプします。なお、egg-convert-regionは、旧Eggのhenkan-regionに相当する関数です。

5 今後の予定

最終的には、Egg V4はGNU Emacs 20に組込まれる予定になっていますが、スケジュールについては私もよく知りません(ごめんなさい)。現在はβテストの準備を行なっているところですので、早く

使ってみたい方やβテストへ参加してみたい方は、開発者&テスターのメーリングリスト tamago ML (tamago@m17n.org; M17N = Multilingualization) へ入会なさってください。入会手順は、

メール本文

subscribe 名前 を tamago-ctl@m17n.org 宛に送る

メール本文

tamago-admin@m17n.org から **subscribe 確認番号 名前** が届く

メール本文

subscribe 確認番号 名前 を tamago-ctl@m17n.org 宛に送り返す

となっています。tamago ML は登録されたメールアドレスからの投稿しか受け付けませんので、投稿専用の別のメールアドレスを登録したい場合は、そのアドレスの登録後に メール本文 **skip** を tamago-ctl@m17n.org へ送ってください。また、Tamago Project のホームページ

<http://www.m17n.org/tamago/>

も御覧になっていて下さい。

6 謝辞

Egg V4 の開発に参加する機会を与えて頂いた g新部@NRI氏 (Egg V4 開発の初代リーダー)、多くの助言を頂いた電総研 Mule グループの戸村氏 (たまごの原作者、Egg V4 開発の現リーダー) 及び半田氏 (Mule の設計者)、そして Egg V4 の開発/テストに参加されている方々に感謝の意を表します。