

Utf82T_EX 利用の手引き

Ver.0906

安田 功

isao@yasuda.homeip.net

2009年 6月 28日

carpe diem. 今日という日を摘め.

Hor. ホラーティウス Carm. I, 11, 8.

目次

1	はじめに	3
2	開発の背景	3
3	機能仕様	4
3.1	機能概要	4
3.2	Utf82T _E X 変換テーブル	10
3.3	機能上の注意事項	10
4	利用上の前提条件	11
5	インストール	12
5.1	UNIX	12
5.2	Windows	15
6	利用の手引	17
6.1	utf82tex コマンドライン書式	17
6.2	Emacs Elisp	18
6.3	拡張変換テーブル	20
6.4	秀丸エディタ	21
6.5	Jedit X	22
6.6	T _E XShop	22
7	適用例	23

7.1	TeX 文書の準備	23
7.2	ヨーロッパ・ラテン文字言語	24
7.3	キリル文字言語	24
7.4	教会スラヴ語	24
7.5	ギリシア語	31
7.6	タイ語	33
7.7	ヴェトナム語	35
7.8	Unicode 漢字, 異体字, 機種依存文字	36
7.9	中韓文字	36
7.10	Emacs 環境	39
8	ユーティリティ	40
8.1	ulatex	40
8.2	genchrtbl	40
9	その他	41
9.1	利用条件	41
9.2	謝辞	41
9.3	misima	41
9.4	変更履歴	43
	参考文献	43

1 はじめに

本文書は Utf82T_EX とその関連ツールの機能仕様、インストール、利用方法について示す。

Utf82T_EX は、UTF-8 文字テキストを pT_EX で利用可能な形式に変換するプログラムである。pT_EX で処理することを想定しつつ、T_EX 文書を主要ヨーロッパ言語の本来の姿で記述することを目的としている。西欧語、ロシア・東欧語、北欧語、教会スラヴ語を含むほとんどのキリル諸言語、複式アクセント文字からなる古典ギリシア語、中国語、韓国語をカバーしている。

Utf82T_EX は UNIX・Windows コマンドラインから直接 `utf82tex` コマンドを発行して多言語テキストを変換することを基本インタフェースとする。またいくつかのエディタ・アプリケーションのバックエンドとして利用するための追加ユーティリティを備えており、GNU Emacs・Windows Meadow、Windows 秀丸エディタ、Mac OS X Jedit X、T_EXShop において編集時の多言語テキストの変換操作をサポートしている。

パッケージは Perl 言語で記述された本体プログラムと、これを GNU Emacs 上で操作するための Elisp プログラム、秀丸エディタ用マクロ・Java クラスライブラリ、Jedit X・T_EXShop 用 AppleScript、その他ユーティリティからなる。パッケージ最新版は <http://yasuda.homeip.net/dl/dl.html> 「T_EX 関連ツール」からダウンロードできる。

2 開発の背景

日本語混在の多言語 T_EX 文書を作成する場合、いちばん困るのはロシア語、フランス語などの非 ASCII コードの言語をすべて ASCII コードで表現せざるをえないことである。“Я люблю вас.” というロシア語は、ローマントランスクリプションによって、“YA lyublyu vas.” などと記述する。同様にフランス語 “Ça, déjeuner!” は “\c Ca, d\’ejeunons!” と。これでは面倒だし、直感的でないがゆえに、スペルミスなど誤りも多くなると考えるのは筆者だけではないはずだ。

多言語 T_EX 文書を多言語コード Unicode で記述する試みに Omega (Lambda)、X_qT_EX がある。しかしながらこれらは、Babel その他の言語パッケージのフォント/ハイフネーション環境との互換性、和文組版品質において、筆者の知る範囲では、まだ完成の域には達していないようである。L^AT_EX 2_ε の枠組においては、unicode パッケージ、CJK パッケージ、Babel パッケージを併用する和文混在多言語組版方法もあり、これ自体たいへん高機能ではあるけれども、設定が面倒でもあり、縦組の品質がいまひとつであるなど和文の扱いに制限があるようだ。ある程度広範囲の外国語が混在した T_EX 文書を欧文正書法に準拠しつつ作成したいけれども、なにより和文がきちんと組めないと困る。

このような状況において、多言語を直接 T_EX 原稿に記述し、なおかつ pT_EX 資産を引き継いだ形で多言語 T_EX を実現するという課題解決のひとつとして Utf82T_EX は開発された。

最近、Unicode 対応版 pT_EX について個人による拡張に着目すべき試みがある。土村展之氏は 2006/8/27 以降の ptetex3 で pT_EX UTF-8 対応パッチを公開している。^{*1} UTF-8 文書をコンパイルするには `--kanji=utf8` オプション付きで `platex` コマンドを起動する。`\usepackage[utf8]{inputenc}` をプリアンブルで指定しておけば、ドイツ語、フランス語などのメジャーなラテン文字言語なら UTF-8 直接入力でも処理できるようになった。

^{*1} <http://www.nn.iiij4u.or.jp/~tutimura/tex/ptetex.html>

また、2007/3/8 ttk 氏によって up \TeX /up \LaTeX が公開された。^{*2} これは土村氏 ptetex3 の Unicode 対応をさらに推し進めて、p \TeX 内部の Unicode 化を果たしたディストリビューションである。 `\kcatcode` プリミティブの拡張、`\disablecjktoken` プリミティブの追加によって、欧文と和文 (漢字) の制御を試みている。 up \TeX /up \LaTeX は、ptetex3 では和文扱いとなってしまうロシア語やギリシア語についても、フォント・正書法に関しきちんと欧文として処理し分けられるようにしただけでなく、CJK Extension にある文字種をサポートしており日中韓の文字空間を大幅に拡張している。

こうした p \TeX そのものの Unicode 対応の進展とともに、Utf82 \TeX プリプロセッサの意義は薄らぎつつある。けれどもタイ語など非 Unicode エンコーディングで組版しなければならない言語パッケージもまだ存在しているし、上記 Unicode 対応の試みが UNIX 以外のプラットフォームにも広がり、定着するにはもう少し時間がかかりそうである。アスキー p \TeX をベースとした過去の資産をこれまで通りに利用できる、という意味において Utf82 \TeX はまだ余命を保っているといえるかも知れない。

3 機能仕様

3.1 機能概要

Utf82 \TeX は UTF-8 でエンコードされたファイルを標準入力から読み込み、p \LaTeX 2 ϵ で処理可能な文字、コントロールシーケンス (命令) に変換して標準出力に書き出す、言わばプリプロセッサである。以下の機能説明で「**[標準]**」とした機能はオプション指定なしに実現されるものであることを示す。「**[-x オプション]**」としたものは“-x”実行オプション指定で働くものである。

3.1.1 標準変換機能

Latin-1, Latin-2 のラテン文字及び Unicode Cyrillic U+0400–04FF^{*3}, Greek and Coptic U+0370–03FF^{*4}, Greek Extended U+1F00–1FFF^{*5} で定義されているキリル文字, ギリシア文字を, ローマントランスクリプション, もしくは \TeX シンボル (テキスト) 命令に変換する。コプト文字, 一部のアクセント単体・記号類はサポートしていないものもある。システム標準変換テーブル (`euro-tbl`) に依って変換する。**[標準]**

3.1.2 ドイツ語変換機能

ラテンアクセント付き文字のうち, `ÄäÖöÜüß` をそれぞれ "A" "a" "O" "o" "U" "u" "s" のように \ なしで出力する。**[-g オプション]**

3.1.3 Bib \TeX 変換機能

Bib \TeX ではラテンアクセント付き文字 `á` を `\'a` と記述するとエラーとなるが, ブレースで囲むとこれを回避できる場合がある。本オプションは, ブレースを付加して出力する。例えば `á` を `{\'a}` に変換する。**[-b オプション]**

^{*2} <http://homepage3.nifty.com/ttk/comp/tex/uptex.html>

^{*3} <http://www.unicode.org/charts/PDF/U0400.pdf>

^{*4} <http://www.unicode.org/charts/PDF/U0370.pdf>

^{*5} <http://www.unicode.org/charts/PDF/U1F00.pdf>

3.1.4 Unicode CJK 漢字変換機能

JIS 第一・第二水準に含まれない Unicode CJK 統合漢字を、日中韓 OTF シーケンス変換（項 3.1.9・7 頁，項 3.1.10・7 頁参照）の指定に応じて、`\UTFx{UCS2 十六進数形式}` ないし `\CIDx{CID 番号}` 形式に変換して出力する。`\UTFx\CIDx` は齋藤修三郎氏による OTF パッケージの提供するコントロールシーケンスである（本稿では以下「OTF シーケンス」と称する場合がある）。Unicode CJK 統合漢字は U+20000 以上のサロゲートペア領域にある拡張 B 漢字もサポートしている。[標準]

変換に際して `Utf82TeX` は、拡張変換テーブル、システム標準変換テーブル (`euro-tbl`)、CJK 変換テーブル (`cjk-tbl`) の順にテーブルを参照する。CJK 変換テーブルは CJK 統合漢字について、UCS コードに対して日本語、中国語簡体字、中国語繁体字、韓国語それぞれの CID 番号を保持している。本テーブルの構築において、文献 [1, 2, 3, 4]、並びに Adobe Systems Incorporated. 提供の CMap ファイル (`UniJIS-UTF16-H`, `UniGB-UTF16-H`, `UniCNS-UTF16-H`, `UniKS-UTF16-H`)、角藤亮氏作成コード変換プログラム `utf8toutf.c` を参照した。

Unicode エリアに対する `Utf82TeX` の CJK 変換テーブルに基づいた変換仕様について、表 1 (6 頁) に示す。

3.1.5 ハングル変換機能

ハングルと判断できる Unicode 文字を `\UTFK{UCS2 十六進数形式}` 命令に変換する。表 1 (6 頁) 参照。[標準]

3.1.6 特殊異体字変換機能

“▼”を前置することによって、特殊異体字を出力できる場合がある。たとえば「▼吉」とすると、「吉」（吉野家の土吉）に変換する（この「つちよし」は Unicode 拡張 B で定義されたもので、直接この文字が入力されても、`Utf82TeX` は「吉」に変換する）。変換対象文字の場合、置換テキストのみを出力し、“▼”は出力しない。変換はテーブルに依存する。登録文字の一覧を表 2 (7 頁) に示す。登録基準はとくにない。筆者の選定に過ぎない。利用者の必要に応じて拡張変換テーブルとして異体字登録を行うことを想定している。

システム標準変換テーブルの定義を図 1 (8 頁) に示す。`c` に続く文字が変換対象となる文字であり、そのすぐ右が変換後の文字シーケンスである。追加修正はシステム標準変換テーブル (`euro-tbl`)、もしくは拡張変換テーブルに対して、図 1 に示した定義を真似て行えばよい。[標準]

3.1.7 タイ語変換機能

タイ語単語境界分割プログラム `swath` 及び、コード変換プログラム `iconv` を使って、UTF-8 タイ語テキストを TIS-620 タイ語文字コードに変換した上、単語境界 (`{\wbr}`) を挿入し、「^^十六進数形式」に変換する。例を図 2 (9 頁) に示す。「^^十六進数形式」とすることによって、本来 TIS-620 でテキストを記述しなければならないタイ語 `ThaiLaTeX` が `pTeX` でも処理できるようになる^{*6}。[`-t` オプション]

^{*6} `swath`, `ThaiLaTeX` の活用については「CJK, Unicode パッケージ, ロシア語多書体の利用」<http://yasuda.homeip.net/tex/cjk.html> を参照。

表 1 Unicode 領域変換仕様

コード範囲	文字ブロック	変換仕様
1100 ~ 11FF	Hangul Jamo (ハングル字母)	<code>\UTFK{十六進数形式}</code> 形式に変換.
2000 ~ 2FFF	General Punctuation (一般句読点) ほか各種記号	<code>\UTFx</code> 形式に変換. ただし日本語 JISX 0208 記号はそのまま出力.
3000 ~ 303F	CJK Symbols and Punctuation (CJK の記号及び句読点)	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換. ただし日本語 JISX 0208 句読点, 記号はそのまま出力する.
3040 ~ 30FF	Hiragana, Katakana	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式に変換, ヒットしなければそのまま出力
3100 ~ 33FF	Bopomofo, Hangul Compatibility Jamo, Kanbun ほか CJK 各種記号, 字母	ハングルは <code>\UTFK</code> 形式に変換. テーブルヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換.
3400 ~ 9FFF	CJK Unified Ideographs (CJK 統合漢字, 拡張 A)	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式に変換, ヒットしなければ日本語の場合そのまま出力. 中韓指定時は <code>\UTFx</code> 形式に変換.
AC00 ~ D7A3	Hangul Syllables (ハングル音節文字)	<code>\UTFK</code> 形式に変換.
E800 ~ FAFB	Private Use Area (私用領域), CJK Compatibility Ideographs (CJK 互換漢字)	テーブルヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換.
FE20 ~ FE2F	Combining Half Marks (半記号)	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換.
FE30 ~ FE6F	CJK Compatibility Forms (CJK 互換形) ほか	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換.
FF00 ~ FFEF	Halfwidth and Fullwidth Forms (半角・全角形)	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式, ヒットしなければ <code>\UTFx</code> 形式に変換. ただし日本語 JISX 0208 句読点, 記号はそのまま出力する.
20000 ~	CJK Unified Ideographs (CJK 統合漢字拡張 B)	CJK 変換テーブルでヒットすれば <code>\CIDx</code> 形式に変換, ヒットしなければ日本語の場合そのまま出力. 中韓指定時は <code>\UTFx</code> 形式に変換.
0080 ~ 1FFF	Latin-1 Supplement ほか各国語文字	各変換テーブルに基づいて変換. ヒットしない文字はそのまま出力.
0000 ~ 007F	Basic Latin (基本ラテン文字) 上記以外	そのまま出力. <code>\UTFx</code> 形式に変換.

3.1.8 拡張変換テーブル追加機能

Utf82TeX では, 利用者独自の文字変換を定義した拡張変換テーブルを追加できる. `$UTF82TEXRC` の `$EXT_DIR` パラメータに指定したディレクトリ下に利用者作成のテーブルを格納し, `-x` オプション付きで実行する.

表 2 特殊異体字変換

入力	出力	入力	出力	入力	出力	入力	出力	入力	出力	入力	出力
▼吉	吉	▼高	高	▼崎	崎	▼碕	碕	▼浜	濱	▼濱	濱
▼角	角	▼第	才	▼間	間	▼葛	葛	▼芦	芦	▼淫	姪
▼群	羣	▼窓	窗	▼翻	翻	▼餅	餅	▼略	畧	▼隣	鄰
▼連	聯	▼欲	慾	▼糾	糾	▼萌	萌				

パッケージには大友康寛氏^{*7}より提供されたグルジア語、ヴェトナム語及び機種依存文字の変換テーブルを同梱している。 [-x オプション]

3.1.9 日中韓 OTF シーケンス変換機能 (オプション指定)

CJK 統合漢字を、オプション指定に応じて `\UTFC`・`\CIDC` (中国語簡体字)、`\UTFT`・`\CIDT` (中国語繁体字)、または `\UTFK`・`\CIDK` (韓国語) の OTF シーケンスに変換する。日中韓テキストを同一フォントで処理すると、Unicode CJK 統合漢字仕様で同一コードポイントに割り当てられた文字で、かつ各国語表記で微妙に字体の異なるものが、同一文字に包摂されて組版されてしまう。位置の異なる句読点も一律になってしまう。本オプション変換を行うと、OTF 機能の恩恵に与って、日中韓の字体・句読点を出力し分けることができる。句読点は「.,.,」を対象とし、変換処理で直前に“`\penalty10000`”を挿入する。 [-c オプション]

3.1.10 日中韓 OTF シーケンス変換機能 (制御タグ指定)

`-c オプション`では CJK 漢字全般に対し指定変換がひとしなみに適用されてしまう。これに対し、対象テキストをタグで明示することによって適用範囲/変換種別を制御することができる。`<utf82tex_x>` タグ以降の漢字を `\UTFx{UCS2 十六進数}` に変換する。すなわち、`x` の値が `c`, `t`, `k` によってそれぞれ `\UTFC` (中国語簡体字)、`\UTFT` (中国語繁体字)、`\UTFK` (韓国語) OTF シーケンスに変換する。`</utf82tex>` タグによって元の変換規則に戻る。`<utf82tex_x>`, `</utf82tex>` タグは出力されない。

仮に `-c オプション`で `K` (韓国語) が指定されていても `<utf82tex_c>` タグ以降 `</utf82tex>` タグもしくは別の `<utf82tex_x>` タグまでは中国語簡体字として扱う。和文の間に中国語、韓国語が混在する場合、本タグでマーキングしておく、オプション如何によらず適切に `\UTFx` シーケンスで出力し分けることができる。例を図 3 (9 頁) に示す。 [標準]

3.1.11 十六進数形式変換機能

コードポイント U+0080 超エリアの UTF-8 文字を「`^^十六進数形式`」に変換する。テーブル変換対象文字、CJK 統合漢字、仮名、ハングル及び JIS 句読点は、本変換の対象には含まれない。pTeX では、8 ビットキャラクターの入力ではエラーとなっても、この形式だと正常に処理できる場合がある。 [-h オプション]

^{*7} <http://www.klavis.info/>

```

c 吉 \CID{13706}      # tsuchi-yoshi
c 高 \CID{8705}      # hashigo-taka
c 崎 \CID{14290}     # tatsu-saki
c 碕 \CID{17009}     # tatsu-saki
c 浜 \CID{8531}      # mayu-hama
c 濱 \CID{8531}      # mayu-hama
c 角 \CID{13682}     # sagari-tuno
c 第 \CID{13910}     # dai-zokuji
c 間 \CID{13693}     # tsuki-ken
c 葛 \CID{7652}      # かつ飾区のみず
c 芦 \CID{7961}      # 芦
c 淫 \CID{4580}      # 姪
c 群 羣             #
c 窓 \CID{5932}      # 窗
c 翻 翻             #
c 餅 餅             #
c 略 畧             #
c 隣 鄰             #
c 連 \CID{13517}     # 聯
c 欲 慾             #
c 糾 糾             #
c 萌 萌             #

```

図 1 特殊異体字定義

3.1.12 キリル文字 T2A 命令変換機能 (オプション指定)

T2A フォントエンコーディングで定義されたキリル文字を、T2A シンボル命令で出力する。pTeX ではもっぱら OT2 フォントエンコーディングを用いるが、PSCyr キリル Type1 フォントパッケージの書体を利用したい、というような T2A で処理しなければならない場合に利用する。変換の定義は T2A 変換テーブル (`t2a-tbl`) に記述されている。 [`-r オプション`]

3.1.13 キリル文字 T2A 命令変換機能 (制御タグ指定)

`<utf82tex_r>` 開始タグと `</utf82tex_r>` 終了タグで囲んだテキスト範囲を選択的に T2A シンボル命令に変換する。タグは変換結果に出力されない。 [`標準`]

[変換前 (UTF-8 タイ語)]

เป็นมนุษย์

[変換後 (TIS-620 十六進数)]

^^e0^^bb^^e7^^b9{\wbr}^^c1^^b9^^d8^^c9^^c2^^ec{\wbr}

図 2 タイ語文字変換

[変換前]

%<utf82tex_k> 韓国語

平骨練。

%<utf82tex_c> 中国語簡体字

平骨練。

%<utf82tex_t> 中国語繁體字

平骨練。

%</utf82tex> 和文に復帰

平骨練。森鷗外

[変換後]

\UTFK{5e73}\UTFK{9aa8}\UTFK{7df4}\penalty10000\UTFK{3002}

\UTFC{5e73}\UTFC{9aa8}\UTFC{7df4}\penalty10000\UTFC{3002}

\UTFT{5e73}\UTFT{9aa8}\UTFT{7df4}\penalty10000\UTFT{3002}

平骨練。森\UTFM{9dd7}外

[タイプセット結果]

平骨練。

平骨練。

平骨練。

平骨練。森鷗外

図 3 日中韓制御タグ

3.1.14 OldSlav 教会スラヴ語変換機能 (オプション指定)

\LaTeX (Slav \TeX) 古代教会スラヴ語パッケージの記法による教会スラヴ語テキストを OldSlav 文字シンボルに変換する。変換の定義は教会スラヴ語変換テーブル (`ocs-tbl`) に記述されている*8。 [`-s オプション`]

*8 \LaTeX , OldSlav 教会スラヴ語パッケージについては「古代教会スラヴ語 \LaTeX 利用の手引き」<http://yasuda.homeip.net/oldslav/oldslavdoc.pdf> 及び「 \TeX での古代教会スラヴ語の利用について」<http://yasuda.homeip.net/oldslav/oldslavtex.html> を参照。

3.1.15 OldSlav 教会スラヴ語変換機能 (制御タグ指定)

`<utf82tex_s>` 開始タグと `</utf82tex_s>` 終了タグで囲んだテキスト範囲を選択的に OldSlav 教会スラヴ語シンボルに変換する。タグは変換結果に出力されない。 **[標準]**

3.2 Utf82TeX 変換テーブル

Utf82TeX の外国語文字変換機能は UTF-8 文字入力文字 \mathring{A} に対して、変換テーブル定義内容に従って、“`\A`” に置き換えることを基本とする。変換後のシーケンスの処理は、 \LaTeX のフォントエンコーディングに依存する。変換テーブルにおける文字変換仕様は表 3 (10 頁) の原則に従っている。

表 3 変換テーブル仕様

文字種別		変換テーブル仕様
ラテン文字		T1 エンコーディング
キリル文字	ロシア語 (旧正書法含む)	可能な限り OT2 エンコーディング (ローマントランスクリプション)。T2A 変換指定時は T2A フォントエンコーディングの文字命令に準拠。
	ウクライナ語ほか現代スラヴ語 その他キリル諸語	同上。 T2A, T2B, T2C エンコーディング
教会スラヴ文字	OldSlav 変換指定 上記以外	OldSlav (LST) エンコーディング T2D エンコーディング。
ギリシア文字		LGR エンコーディング (一部 Teubner パッケージ記法あり)

\LaTeX OT2 キリル・フォントエンコーディングにおけるローマントランスリテレーションの問題は、キリル文字と一対一対応でないがゆえの曖昧さである。たとえば、“`tc`” に対し“`ts`” を入力すると、通常は LH フォントのリガチャの定義に従って“`т`” で組版されてしまう。Utf82TeX ではこの問題への対応として、ロシア語正書法に鑑み、以下 2 点の変換仕様を試みに採用している。

- “`tc`” は、“`c`” または “`C`” の直前に “`t`” または “`T`” が出現するとき、“`t\~s`” (“`t`”, “`s`” はそのときのケースにより大文字となる) に変換する。“`\~`” はハイフネーション許可を示す。
- “`cx`” は、“`x`” または “`X`” の直前に “`c`” または “`C`” が出現するとき、“`s\textcompwordmark{h}`” (“`s`”, “`h`” はそのときのケースにより大文字となる) に変換する。“`\textcompwordmark`” はリガチャ/ハイフネーション禁止を示す。

3.3 機能上の注意事項

- 変換対象外の文字についてはそのまま出力する。主に日本語仮名は UTF-8 で出力されるわけで、Utf82TeX 実行後、 \LaTeX 2_ε コンパイル実行の前に、さらに `iconv`, `nkf` 等のコード変換ツールを

利用して日本語 EUC, SJIS, JIS コードに変換しておく必要がある。

- Utf82T_EX で処理するのはあくまで文字に限定されるので、各種言語を T_EX で処理するために必要な指定（たとえば Babel 言語オプション指定など）は利用者でコーディングしておくことに注意されたい。
- ThaiT_EX, OTF など利用機能の前提となる T_EX パッケージはあらかじめ導入しておき、T_EX 文書にもしかるべく指定しなければならない。

4 利用上の前提条件

Utf82T_EX を使用するために必要な環境、サポート機能に付随する周辺ソフトウェアについて示す。

- Utf82T_EX 本体は実行環境として Perl 5.8.0 以降が必要である。“`perl -version`”とコマンドを投入してお手持ちのバージョンを確認いただきたい。筆者は FreeBSD 7-RELEASE, Mac OS X 10.4 Tiger にて開発/動作試験を行っている。Linux などの他の UNIX 系 OS でも動作するはずである（未確認）。Windows 2000 SP.4 (ActivePerl) で動作確認済みである。Windows 98/Me はタイ語オプションをサポートしていない。
- タイ語処理機能（`-t` 及び `-T` オプション）を利用する場合は、`swath`^{*9} 及び `iconv` の各外部コマンドがあらかじめインストールされていなければならない。
- Utf82T_EX は UTF-8 エンコーディングテキストを入力前提とする。したがってテキストの編集には UTF-8 が扱えるエディタが必要である。Emacs Utf82T_EX ドライバ Elisp を利用するには、当然ながら GNU Emacs が前提となる。
- Emacs 20 または 21 で教会スラヴ語、キリル言語（ISO 8859-5 で定義されていない文字）、古典ギリシア語を表示・編集するには、別途 Unicode サポートパッケージが必要である。Emacs 20.7 では Mule-UCS または Otfried Cheong 氏作成の `oc-unicode` パッケージ^{*10} が利用できる。Emacs 21 では Mule-UCS のみである。Emacs 22 は標準で Unicode をサポートしている。
- Emacs 用 Unicode フォントについては、電子書体オープンラボが配布する `/efont/`^{*11}, Markus Kuhn 氏による `ucsfonts` フォント^{*12} 等を組み込んでおく。
- Emacs では Leim パッケージによって多国語インプットメソッドがサポートされている。Unicode キリル言語、古典ギリシア語について筆者は Quail インプットメソッドを提供している。利用方法は「Emacs スラヴ語/古典ギリシア語汎用インプットメソッド」^{*13} を参照。
- Emacs スラヴ語/古典ギリシア語汎用インプットメソッド, Utf82T_EX Elisp は FreeBSD Emacs 20, 21 及び Mac OS X Tiger X11-Emacs 22.0.50 で開発・試験した。Windows Meadow でも利用可能である。Mac OS X では X11 用にビルドした Emacs であれば動作する。それぞれの Unicode 環境については、「Emacs 21 Unicode 環境」^{*14}, 「Windows Meadow 2.10 多言語環境の設定」^{*15}, 「Mac OS X

^{*9} <http://www.cs.cmu.edu/~paisarn/software.html>

^{*10} <http://www.cs.uu.nl/~otfried/Mule/>

^{*11} <http://openlab.ring.gr.jp/efont/unicode/index.html.ja>

^{*12} <http://www.cl.cam.ac.uk/~mgk25/ucs-fonts.html>

^{*13} <http://yasuda.homeip.net/rus2/emacs-im.html>

^{*14} <http://yasuda.homeip.net/rus2/emacs21.html>

^{*15} <http://yasuda.homeip.net/rus2/meadow.html>

Tiger, X11-Emacs 22.0.50」*16 を参照。

- Mac OS X Carbon-Emacs では、BDF 多言語フォントが使えない関係で、添付 Elisp を利用することができない。X11-Emacs でご活用いただきたい。
- Windows Meadow ではバージョンによって `utf82tex.el` がうまく動作しない場合がある。その場合はコマンドプロンプトから実行する。Emacs 22 ベースの Meadow 3.00 では動作するようであるが、筆者は Meadow では十分な確認ができていない。
- 秀丸エディタについてはバージョン 6.09 にて試験を行っている。Windows のバージョンは XP もしくは 2000 である。Utf82TeX 用秀丸マクロは、Unicode テキストを取り扱う関係でクリップボードインタフェースを使用しており、クリップボードデータを Utf82TeX によって変換し、そのテキストを編集画面にペーストする仕組みになっている。このときクリップボードデータの入出力に misima 旧字・旧仮名遣い変換プログラムの Java クラスライブラリを利用しており、従って秀丸エディタ用ツールは Java 5.0 以上の実行環境 (JRE 5.0 ないし JDK 5.0 以上) を必要とする。
- Jedit X 及び T_EXShop 用の AppleScript は Mac OS X 10.4 Tiger で確認している。Jedit X エディタはバージョン 1.47 にて、T_EXShop エディタはバージョン 2.14 にて、それぞれ試験を行っている。

5 インストール

5.1 UNIX

UNIX (Mac OS X を含む) での導入手順を説明する。%, # はそれぞれ Tcsh の一般ユーザ, スーパーユーザのコマンドプロンプトを示す。

1. パッケージの入手と展開

```
% wget -nH -nd http://yasuda.homeip.net/archives/\
  utf82tex-0906.tar.gz
% gunzip -c utf82tex-0906.tar.gz | tar xvf -
% cd utf82tex-0906
```

2. インストールパスの修正

パッケージにはインストール用 `Makefile` を添付している。以下に示す格納パスなどの変数値を適宜修正する。Emacs 21 のユーザは 2006.12 の時点では `EMACSVAR` を `21.3` とすればよいはずである。また Emacs 22 (Emacs-cvs 版) では `22.0.90` などとする。Emacs 20.7 で `oc-unicode` を利用する場合は、`OCUNICODE` に `1` を設定する。`PREFIX`, `UTF82TEXDIR` は変更しないことをお勧めする。

```
PREFIX          = /usr/local
EMACS           = $(PREFIX)/bin/emacs
```

*16 <http://yasuda.homeip.net/rus2/mac-emacs22.html>

```
OCUNICODE      = 0
EMACSVER       = 22.0.50
EMACSPATH      = $(PREFIX)/share/emacs
LEIMPATH       = $(EMACSPATH)/$(EMACSVER)/leim
UTF82TEXDIR    = $(PREFIX)/etc/utf82tex
UTF82TEXPATH   = $(PREFIX)/bin
UTF82TEXELPATH = $(EMACSPATH)/$(EMACSVER)/site-lisp
TEXDIR         = $(PREFIX)/teTeX/share/texmf-local
```

3. GNU make によるインストール

スーパーユーザ権限で GNU make を実行する。FreeBSD では `gmake` コマンドを用いる。`utf82tex`, `utf82tex.elc` が所定のパスに格納される。

```
# make utf82tex-install
```

4. 手動によるインストール

添付 `Makefile` によらず、手動でインストールする手順を示す。

```
% cd utf82tex-0906
% su -
[Utf82TeX 用ディレクトリ作成]
# mkdir -p /usr/local/etc/utf82tex
[各種ファイルのコピー]
# cp -p config/utf82texrc /usr/local/etc/utf82tex
# cp -p tables/*-tbl /usr/local/etc/utf82tex
# cp -p -R contrib /usr/local/etc/utf82tex
[utf82tex モジュールのインストール]
# install -v -m 755 modules/utf82tex /usr/local/bin
# install -v -m 755 modules/utf8cat /usr/local/bin
# install -v -m 755 utils/ulatex /usr/local/bin
# install -v -m 755 utils/genchrtbl /usr/local/bin
[utf82tex.el のバイトコンパイル]
# cd elisps
# emacs -batch -f batch-byte-compile utf82tex.el
[utf82tex.elc の Emacs ロードパスへの格納]
# cp -p utf82tex.elc /usr/local/share/emacs/site-lisp
# exit
```

```
% rehash
```

5. UTF82TEXRC 環境変数の設定

環境変数 `UTF82TEXRC` にコンフィギュレーションファイルの絶対パスを設定する。ログインシェルの初期設定ファイル (tcsh なら `.tcshrc`) に記述しておくといよい。この環境変数がセットされていない場合、またコマンドラインオプション `-f` でファイル指定しない限りにおいて、`/usr/local/etc/utf82tex/utf82texrc` を仮定する。本設定は付属 `Makefile` の標準パスでインストールした場合不要である。

```
tcsh:
    setenv UTF82TEXRC /usr/local/etc/utf82tex/utf82texrc
bash:
    export UTF82TEXRC=/usr/local/etc/utf82tex/utf82texrc
```

6. コンフィギュレーションファイルの編集

パス情報システム標準値を変更したいなら、インストール環境に応じてコンフィギュレーションファイル設定値 (式の右辺) を修正する。記述は Perl の文法に準拠する必要がある。とくにタイ語関連外部プログラム `$SWATH`, `$ICONV` のパス, `$SWOPT` オプションを確認する。UNIX 系 OS で、添付 `Makefile` の標準設定で導入したなら、コンフィギュレーションファイルの編集は必要ないと思う。

```
# iconv プログラムのファイル名を絶対パスで指定する。
$ICONV = "/usr/local/bin/iconv";
# iconv プログラムのオプションを指定する。(変更不可)
$ICOPT = "-c -f UTF-8 -t TIS620";
$ICOPTR = "-c -f TIS620 -t UTF-8";
# swath プログラムのファイル名を絶対パスで指定する。
$SWATH = "/usr/local/bin/swath";
# swath プログラムのオプションを指定する。(原則変更不可)
# ただし単語分割アルゴリズムを追加してもよい(以下は標準)
$SWOPT = "-f latex";
# $ECHO echo (for Unix)
$ECHO = "/bin/echo";
# システム標準変換テーブルを絶対パスで指定する。
$EUR_TBL = "/usr/local/etc/utf82tex/euro-tbl";
# CJK 変換テーブルを絶対パスで指定する。
$CJK_TBL = "/usr/local/etc/utf82tex/cjk-tbl";
# キリル T2A 変換テーブルを絶対パスで指定する。
```

```
$T2A_TBL = "/usr/local/etc/utf82tex/t2a-tbl";
# 教会スラヴ語変換テーブルを絶対パスで指定する.
$OCS_TBL = "/usr/local/etc/utf82tex/ocs-tbl";
# 拡張変換テーブル格納ディレクトリ名を絶対パスで指定する.
# 拡張変換テーブルはいくつ登録してもよい.
$EXT_DIR = "/usr/local/etc/utf82tex/contrib";
```

5.2 Windows

`C:¥>` はコマンドプロンプトでの入力を示している. ここでは `Utf82TeX` のみの説明である.

1. パッケージの入手と展開

ブラウザなどで <http://yasuda.homeip.net/archives/utf82tex-0906.zip> をダウンロードし, ワークフォルダ (ここでは `C:¥tmp` とする) に解凍する.

2. インストールパスの修正

パッケージには `Utf82TeX` インストール用バッチファイル `install.bat` を添付している. これをエディタで開いて, 以下に示すインストール先パス情報などの変数値を適宜修正する. `ℒTeX` を `C:¥usr¥local¥bin` にインストールしているなら修正は不要である. 標準では表 4 (16 頁) のとおりインストールされるものとしている.

```
set bindir=C:¥usr¥local¥bin
set screxe=runscr.exe
set utf82texdir=C:¥usr¥local¥etc¥utf82tex
set extdir=C:¥usr¥local¥etc¥utf82tex¥contrib
```

とくに `screxe` 変数値を確認いただきたい. 角藤先生の `W32TeX`^{*17} 関連パッケージ `ConTeXt`^{*18} がインストールされていると, `ℒTeX` プログラムをインストールした実行バイナリフォルダ (通常 `C:¥usr¥local¥bin`) のなかに `runscr.exe` かもしくは `perlexec.exe` が存在するはずである. この名称を `screxe` 変数値に指定する.

3. バッチコマンドによるインストール

`install.bat` を実行する. 次のようにコマンドプロンプトから指示してもよいし, エクスプローラでこのファイルをダブルクリックしてもよい.

```
C:¥tmp¥utf82tex-0906>install
```

*17 <http://www.fsci.fuk.kindai.ac.jp/kakuto/win32-ptex/web2c75.html>

*18 <ftp://akagi.ms.u-tokyo.ac.jp/pub/TeX/win32/context.tar.gz>

表 4 Windows インストールパス標準値

データ	パス標準値
utf82tex	C:%usr%local%bin%utf82tex
utf8cat	C:%usr%local%bin%utf8cat
screxe	C:%usr%local%bin%runscr.exe
iconv	C:%usr%local%bin%iconv.exe
swath	C:%usr%local%bin%swath.exe
swath データ	C:%usr%local%bin%data
utf82texrc	C:%usr%local%etc%utf82tex%utf82texrc
utf82tex テーブル	C:%usr%local%etc%utf82tex
utf82tex 拡張テーブル	C:%usr%local%etc%utf82tex%contrib

4. UTF82TEXRC 環境変数の設定

環境変数 `UTF82TEXRC` にコンフィギュレーションファイルの絶対パスを設定する。Windows ではシステムプロパティで以下の対を登録する。

```

キー: UTF82TEXRC
値  : C:%usr%local%etc%utf82tex%utf82texrc
    
```

5. コンフィギュレーションファイルの編集

エディタで `C:%usr%local%etc%utf82tex%utf82texrc` を必要に応じて編集する。雛形は `config%utf82texrc-win` にある。標準値を変更する場合、インストール環境に応じてコンフィギュレーションファイル設定値（式の右辺）を修正する。記述は Perl の文法に準拠する必要がある。とくにタイ語関連外部プログラム `$SWATH`, `$ICONV` のパス, `$SWOPT` オプションを確認する。実行可能を示す拡張子 (exe, com) をもつファイル名でなければならない。

```

# iconv プログラムのファイル名を絶対パスで指定する。
$ICONV = "C:/usr/local/bin/iconv.exe";
# iconv プログラムのオプションを指定する。(変更不可)
$ICOPT = "-c -f UTF-8 -t TIS620";
$ICOPTR = "-c -f TIS620 -t UTF-8";
# swath プログラムのファイル名を絶対パスで指定する。
$SWATH = "C:/usr/local/bin/swath.exe";
# swath プログラムのオプションを指定する。
# Windows は単語分割アルゴリズム追加必須 (以下は標準)
$SWOPT = "-f latex -d C:/usr/local/bin/data";
# $CAT utf8cat for Windows only
    
```



```

$CAT      = "perl C:/usr/local/bin/utf8cat";
# 文字変換基本テーブルを絶対パスで指定する.
$EUR_TBL  = "C:/usr/local/etc/utf82tex/euro-tbl";
# CJK 変換テーブルを絶対パスで指定する.
$CJK_TBL  = "C:/usr/local/etc/utf82tex/cjk-tbl";
# キリル T2A 変換テーブルを絶対パスで指定する.
$T2A_TBL  = "C:/usr/local/etc/utf82tex/t2a-tbl";
# 教会スラヴ語用 変換テーブルを絶対パスで指定する.
$OCS_TBL  = "C:/usr/local/etc/utf82tex/ocs-tbl";
# 拡張変換テーブル格納ディレクトリ名を絶対パスで指定する.
$EXT_DIR  = "C:/usr/local/etc/utf82tex/contrib";

```

- フォルダセパレータは / を用いることができる。Windows 本来の “¥” でもよいが、これを指定するときは、Perl 文法に準拠して “¥¥” のように二つ指定しなければならないことに注意。
- Windows でタイ語処理を行う場合、`$SWOPT` のうち `swath -d` オプションを必ず指定する。これは `swath` の分割アルゴリズムデータの格納場所である。タイ語処理のときのみ参照される。詳細は `swath` のマニュアルを参照のこと。
- Perl コマンドへのパスを通しておく。環境変数 `PATH` に `perl.exe` が格納されているフォルダ (`C:¥Perl¥bin` など) が設定されていることを確認しておく。

6 利用の手引

Utf82TeX は、コマンドラインでの `utf82tex` コマンド及び GNU Emacs 関数による二つの変換操作をサポートしている。

6.1 utf82tex コマンドライン書式

```
utf82tex [ -r|-s|-g|-b|-t|-T|-x|-h|-q|-u|-c {KCT}|-f "rcfile" ]
```

- [] 内のオプションは任意である。
- A|B は A と B とを選択的もしくは同時に指定可能であることを示す。-tx のような指定もできる。-x を同時に指定すると、拡張変換テーブル、システム標準変換テーブルの順に内容が優先される。-r、-s オプションを指定すると、それぞれ T2A 変換テーブル、教会スラヴ語変換テーブルの定義が優先される。-c オプションを指定すると指定パラメータに従って漢字について中国語簡体字、中国語繁体字もしくは韓国語への変換が優先される。
- { } 内の文字は -c 指定時 K, C, T のいずれかひとつが必須であることを示す。

6.1.1 標準機能 (オプション指定なし)

- キリル文字, ラテン文字, ギリシア文字をローマントランスクリプションまたは $\text{T}_{\text{E}}\text{X}$ シーケンスに変換する.
- JIS 第一・第二水準で定義されていない CJK 統合漢字を $\backslash\text{UTFM}\{\text{UCS2 十六進数}\}$ もしくは $\backslash\text{CID}\{\text{CID 番号}\}$ 形式の OTF シーケンスに変換する. ハングルは $\backslash\text{UTFK}\{\text{UCS2 十六進数}\}$ に変換する.
- $\langle\text{utf82tex}_x\rangle$ タグ以降の CJK 漢字を, x の値 c, t, k によってそれぞれ $\backslash\text{UTFC}$, $\backslash\text{UTFT}$, $\backslash\text{UTFK}$ の OTF シーケンスに変換する. $\langle\text{utf82tex}\rangle$ タグによって元の変換規則に復帰する. タグそのものは出力されない.
- $\langle\text{utf82tex}_r\rangle$ タグ以降のキリル文字を T2A シンボル命令に変換する. $\langle\text{utf82tex}_r\rangle$ タグによって元の変換規則に復帰する. タグそのものは出力されない.
- $\langle\text{utf82tex}_s\rangle$ タグ以降の文字を OldSlav シンボルに変換する. $\langle\text{utf82tex}_s\rangle$ タグによって元の変換規則に復帰する. タグそのものは出力されない.

6.1.2 オプション

コマンドラインで指定できるオプションについて, 表 5 (19 頁) に示す.

6.2 Emacs Elisp

6.2.1 .emacs への記述追加

Emacs の初期設定ファイル (標準では `.emacs`) に以下の設定を行う.

```
;; Utf82TeX.el のロード
(require 'utf82tex)

;; Utf82TeX のインストールパス
(setq utf82tex-path "/usr/local/bin/utf82tex")

;; Utf82TeX rcfile の絶対パス
(setq utf82tex-rc "/usr/local/etc/utf82tex/utf82texrc")

;; Utf82TeX による変換後のバッファ・エンコーディング
;; Emacs で利用できるコーディングシステムのシンボルを記述する.
;; Emacs の TeX モードでエンコーディングを設定していると指定が
;; 有効にならない場合がある.
(setq utf82tex-out-encoding 'utf-8)
```

表 5 Utf82TeX オプション

オプション	機能
-r	キリル文字を T2A のシンボルに変換する。
-s	文字を OldSlav 教会スラヴ語のシンボルに変換する。
-g	文字をドイツ語 7 ビット形式に変換する。
-b	文字を BibTeX でも通るようブレースで囲んで変換・出力する。
-t	タイ語を <code>iconv</code> によって TIS-620 コードに変換し、 <code>swath</code> によって単語分割境界をマーキングし、結果を「 <code>^^十六進数</code> 」形式で出力する。 <code>iconv</code> 、 <code>swath</code> がインストールされていないとエラーになる。
-T	タイ語を <code>iconv</code> によって TIS-620 コードに変換し、 <code>swath</code> によって単語分割境界をマーキングし、結果を UTF-8 で出力する。 <code>iconv</code> 、 <code>swath</code> がインストールされていないとエラーになる。
-x	拡張変換テーブルを使って文字を変換する。拡張変換テーブル用のディレクトリに格納されたすべてのテーブルを読み込み、システム標準変換テーブルにマージする。拡張変換テーブルの定義はシステム標準変換テーブルよりも優先される。複数の拡張変換テーブルの間で定義が重複している場合、その優先度は Perl <code>readdir</code> 関数が返す拡張テーブルエントリの順番に依存する。2006.12 時点ではグルジア語用、ヴェトナム語用、機種依存文字用拡張変換テーブルをパッケージに添付しており、 <code>-x オプション</code> でこれらを読み込むようになっている。
-c	CJK 統合漢字を <code>\UTFx{UCS2 十六進数}</code> もしくは <code>\CIDx{CID 番号}</code> 形式に変換する。本オプションを指定する場合 <code>-c</code> に続けて以下のいずれかを選択しなければならない。
K	韓国語として扱い、 <code>\UTFK・\CIDK</code> 形式に変換する。
C	中国語簡体字として扱い、 <code>\UTFC・\CIDC</code> 形式に変換する。
T	中国語繁体字として扱い、 <code>\UTFt・\CIDT</code> 形式に変換する。
-h	標準機能変換対象外の文字を「 <code>^^十六進数</code> 」形式に変換する。
-f	コンフィギュレーションファイルを指定する。このオプションを省略したときは環境変数 <code>UTF82TEXRC</code> の設定値が参照される。
-q	致命的エラー以外は報告しない。エラーは標準エラー出力に表示する。
-u	簡易ヘルプ、 <code>\$UTF82TEXRC</code> の各種変数値を表示する。

6.2.2 Emacs での変換操作

Emacs 上で UTF-8 TeX 文書作成をしつつ、そのバッファ、リージョンに対し Utf82TeX 変換を指示するための命令 (関数) を表 6 (20 頁) に示す。関数名に含まれる `-region`、`-buffer` はそれぞれリージョン、バッファに対する操作を表している。関数を呼び出すには対象となるリージョン、バッファを選択した状態で、“`M-x 関数名 RET`”を実行する。

表 6 Emacs 命令

関数名	機能
<code>utf82tex-region</code>	指定したリージョンを <code>Utf82TeX</code> 標準機能 (<code>-q</code> オプション付) で変換し、置き替える.
<code>utf82tex-t2a-region</code>	<code>utf82tex -rq</code> に同じ. キリル T2A 変換.
<code>utf82tex-ocs-region</code>	<code>utf82tex -sq</code> に同じ. 教会スラヴ語変換.
<code>utf82tex-thai-region</code>	<code>utf82tex -tq</code> に同じ. タイ語変換.
<code>utf82tex-thai-ucs-region</code>	<code>utf82tex -Tq</code> に同じ. タイ語変換 (UTF-8 出力).
<code>utf82tex-hex-region</code>	<code>utf82tex -htq</code> に同じ. 十六進数変換.
<code>utf82tex-korean-region</code>	<code>utf82tex -q -c K</code> に同じ. 韓国語変換.
<code>utf82tex-chinese-simplified-region</code>	<code>utf82tex -q -c C</code> に同じ. 中国語簡体字変換.
<code>utf82tex-chinese-traditional-region</code>	<code>utf82tex -q -c T</code> に同じ. 中国語繁體字変換.
<code>utf82tex-extend-region</code>	<code>utf82tex -xq</code> に同じ. 拡張テーブル変換.
<code>utf82tex-buffer</code>	当該バッファを <code>Utf82TeX</code> 標準機能で変換し、置き替える.
<code>utf82tex-create-buffer</code>	当該バッファを <code>Utf82TeX</code> 標準機能で変換し、指定したファイル名で新しいバッファを作成し、カレントバッファを切り替える. 同時にファイルを格納する. この時 <code>utf82tex-out-encoding</code> 変数で指定したコーディングシステムに設定する.

6.3 拡張変換テーブル

利用者が独自に変換規則を追加したい場合、定義ファイルを UTF-8 エンコーディングで作成し、拡張変換テーブル用ディレクトリに格納する. 登録テーブル数はいくつでもよい. 拡張変換テーブルのファイル名は OS 仕様に準拠していれば任意である. 書式は以下のとおりである.

```
# コメント行
e 変換前 変換後 # コメント
c 変換前 変換後 # コメント
```

先頭の文字 `e`, `c` は固定である. `e` は基本変換の定義である. `c` は“▼”を前置したときに適用する特殊異体字変換の定義を示す. 要素は空白文字 (スペース, タブ) で区切って記述する. はじめの二つの空白文字列を区切りとして認識する. 二つ目の空白文字列に続く文字列を「**変換後**」として扱う. 三つ目の空白文字列より

も後の文字列に # が存在すると、当該文字列以降を「コメント」として無視する。「変換前」には変換対象としたい一文字の UTF-8 キャラクタを指定する。「変換後」には、「変換前」を置換すべき文字ないし T_EX コントロールシーケンスを記述する。行頭に # を記述すると当該行をコメントとする。コメントは Utf82T_EX の動作には影響しない。

複数の拡張テーブル間で定義が重複すると、あとから読まれたテーブル内容が優先される。入力の順番は Perl `readdir` 関数の仕様に依存する。重複時のオーバーロード内容は標準エラーに出力されるので、定義が上書きされてしまうような場合は、メッセージを確認してファイル名称を変更するなどの対処が必要となる。

6.4 秀丸エディタ

Utf82T_EX は秀丸エディタの編集画面から選択テキストの多言語テキスト変換が可能である。ここではそのための環境設定と使い方を示す。

6.4.1 インストール

秀丸エディタ用ツールは `install.bat` ではインストールされない。別途以下の手順で設置する。

1. Java クラスライブラリの設置

パッケージ `hidemaru` フォルダにある `misima SOAP Java クラスライブラリ misima.jar` を適当な場所にコピーし、その絶対パスを `CLASSPATH` 環境変数に登録する。

2. 秀丸マクロの登録

`hidemaru` フォルダ内にある `*.mac` マクロファイルを秀丸エディタの「マクロ登録」画面から登録する。マクロファイルと Utf82T_EX オプションとの関係は表 7 (21 頁) のとおり。マクロ名、ショートカットはこれに応じて利用者で分かりやすいものに設定する。

表 7 秀丸マクロ一覧

マクロファイル名	変換操作内容	オプション
<code>utf82tex-std.mac</code>	標準変換	<code>-qx</code>
<code>utf82tex-germany.mac</code>	ドイツ語変換	<code>-qgx</code>
<code>utf82tex-t2a.mac</code>	キリル T2A 変換	<code>-qrx</code>
<code>utf82tex-thai.mac</code>	タイ語変換	<code>-qtx</code>
<code>utf82tex-ocs.mac</code>	教会スラヴ語変換	<code>-qsx</code>
<code>utf82tex-chinese-simplified.mac</code>	中国語簡体字変換	<code>-qx -c C</code>
<code>utf82tex-chinese-traditional.mac</code>	中国語繁体字変換	<code>-qx -c T</code>

3. 注意事項

Utf82T_EX 秀丸マクロは、`perl.exe` 及び `utf82tex` Perl モジュールの場所をハードコーディングしている。もし Utf82T_EX のインストールを標準値以外のフォルダに対して実施した場合、マクロ内のパス定義情報が不整合を起こすため、マクロをエディタで開いてパスを修正すること。このときフォルダ・セパレータ記号 (円記号) は以下のとおり必ず二個 (¥¥) 指定すること。

```
$Perl = "C:¥¥Perl¥¥bin¥¥perl.exe";
$Prog = "C:¥¥usr¥¥local¥¥bin¥¥utf82tex";
```

6.4.2 使い方

TeX 原稿において変換したい多言語テキスト範囲をマウスでドラッグして選択する。この状態で、適当な変換オプションにバインドした Utf82TeX 用マクロを実行すればよい。

6.5 Jedit X

Utf82TeX は Mac OS X Jedit X エディタの編集画面から選択テキストの多言語テキスト変換が可能である。ここではそのための環境設定と使い方を示す。

6.5.1 インストール

Jedit X 用ツールは **Makefile** ではインストールされない。別途以下の手順で設置する。

1. ディスクイメージのマウント

パッケージを展開したディレクトリの直下にある **Utf82TeX-MacOSX.dmg** ディスクイメージをダブルクリックしてマウントする。Finder にある“JeditX”フォルダを開くと、各種 AppleScript ファイルが格納されている。

2. AppleScript の登録

Jedit X を起動し、メニューバー右端のリボンアイコンから「スクリプトウィンドウを表示」をクリックし、スクリプトウィンドウを表示させる。次に、ディスクイメージの“JeditX”フォルダ内のスクリプトを全て選択 (**command+a**) し、マウスでこれらをスクリプトウィンドウ内にドラッグする。これで Utf82TeX 用 AppleScript の組込みは完了である。

6.5.2 使い方

TeX 原稿において変換したい多言語テキスト範囲をマウスでドラッグして選択する。この状態で、適当な変換オプションにバインドした Utf82TeX 用 AppleScript を実行すればよい。

6.6 T_EXShop

Utf82TeX は Mac OS X T_EXShop エディタの編集画面から選択テキストの多言語テキスト変換が可能である。ここではそのための環境設定と使い方を示す。

6.6.1 インストール

T_EXShop 用ツールは **Makefile** ではインストールされない。別途以下の手順で設置する。

1. ディスクイメージのマウント

パッケージを展開したディレクトリの直下にある `Utf82TeX-MacOSX.dmg` ディスクイメージをダブルクリックしてマウントする。Finder にある “TeXShop” フォルダを開くと、各種 AppleScript 用のテキストファイルが格納されている。

2. AppleScript の登録

TeXShop を起動し、「マクロ」-「マクロエディタ」を順次クリックしてマクロエディタを表示させる。「新規マクロ」をクリックして「内容:」のエリアに、“TeXShop” フォルダ内の提供スクリプト内容を貼付ける。スクリプト用テキストファイルを開くには TeXShop そのものを使えばよい。「項目名:」にはスクリプトファイル名に応じた名称を利用者で入力する。この操作をスクリプトの数だけ繰り返して、「保存」をクリックすれば登録は完了である。項目名をマウスでドラッグすることによりメニューの並びを調整しておく。

6.6.2 使い方

TeX 原稿において変換したい多言語テキスト範囲をマウスでドラッグして選択する。この状態で、適当な変換オプションにバインドした Utf82TeX 用 AppleScript を実行すればよい。

7 適用例

7.1 TeX 文書の準備

TeX 文書ファイルを UTF-8 エンコーディングで作成・準備する。スラヴ語はキリル文字、ギリシア語は複式アクセントを含めギリシア文字で記述する。フランス語、ドイツ語もその本来の文字で表現する。ラテン文字については、ISO 8859-1 (Latin-1), ISO 8859-2 (Latin-2) で定義されている文字集合の範囲をサポートしているので主な西欧語、北欧語、東欧語がそのままの姿で記述できるはずである。日本語、中国語、韓国語も包摂の問題を置けば、それぞれの言語を表現できる。

本稿であげる例をコンパイルするためには次のような指定をプリアンブルで行っておけばよい。

```
\documentclass[a4paper]{jsarticle}
\usepackage{inputenc}
\usepackage[OT2, T2A, T2B, T2C, T2D, T1]{fontenc}
\usepackage[thai, vietnam, %Babel by ThaiLaTeX, VnTeX
            russian, ukrainian,
            french, german, czech,
            oldchurchslavonic, %OldSlav
            polutonikogreek,
            nippon]{babel}
\usepackage{thswitch}%ThaiLaTeX encoding switch
\usepackage{dbllacnt}%VnTeX
\usepackage{vietnam}%VnTeX
\usepackage{teubner}%Teubner: Classic Greek
```

```
\usepackage[deluxe,expert,multi]{otf}% OTF 和文
```

Babel, unicode, OTF, ヴェトナム語 VnTeX, タイ語 ThaiTeX, 教会スラヴ語 OldSlav (GikTeX), Teubner の各パッケージが必要である。Babel 言語オプションに `nippon` を指定している。`utf82tex-0906/babel/` 下に日本語言語定義 `nippon.ldf` を添付しているので、これを TeX ツリー (`/usr/local/texlive/texmf-local` 配下のディレクトリ階層、以降「ツリー」はディレクトリ階層を示す) にコピーしておけばよい。

7.2 ヨーロッパ・ラテン文字言語

Utf82TeX は国際標準文字集合 ISO 8859-1 及び ISO 8859-2 で定義されているラテン文字の変換を行う。ISO 8859-1 は、アイスランド語、アイルランド語、イタリア語、英語、オランダ語、スウェーデン語、デンマーク語、ドイツ語、ノルウェー語、ポルトガル語、フィンランド語、フェロー語、フランス語で用いられる文字をカバーしている。一方 ISO 8859-2 でサポートしている言語は、アルバニア語、英語、スロバキア語、スロベニア語、セルビア・クロアチア語、チェコ語、ドイツ語、ハンガリア語、ポーランド語、ルーマニア語である。

図 4 (26 頁), 図 5 (27 頁) にラテン文字言語の直接入力例とその L^AT_EX 処理結果を示す。

7.3 キリル文字言語

Unicode で定義されているすべてのキリル文字の変換が可能である。

極力 OT2 フォントエンコーディング・アスキートランスリテレーションで変換するが、T2A, T2B, T2C のシンボル命令で出力する場合もある。Ѡ, ѡ など革命前のロシア語旧正書法文字の変換出力も可能である。T2A, T2B, T2C フォントエンコーディングで記述できるキリル言語の一覧は `TEXMF:doc/latex/base/cyrguide.dvi` (インターネットリソースでは <http://www.tug.org/tex-archive/macros/latex/doc/cyrguide.pdf> などから閲覧できる) を参照のこと。

`-r` オプションもしくは `<utf82tex_r>`, `</utf82tex_r>` 制御タグを指定することにより T2A シンボル命令に変換する。制御タグ使用時はタグ範囲内に限定して変換する。制御タグは TeX が無視するようコメントとして入れておくとよい。

図 6 (28 頁), 図 7 (29 頁) にキリル文字言語の直接入力例とその L^AT_EX 処理結果を示す。

例では、ウクライナ語において T2A 変換を施し、PSCyr Type1 フォントで組む指示を行っている。pTeX においてキリル文字は OT2 アスキートランスリテレーションで入力することが多いと思われるが、UTF-8 キリル文字を直接タイプし、Utf82TeX T2A 変換結果でタイプセットを行うことで、T2A フォントエンコーディングによる PSCyr 等の Type1 フォントが使用できるようになる。

7.4 教会スラヴ語

Unicode にはすべてではないが古スラヴ語／教会スラヴ語の字母が含まれている。Utf82TeX はこれらの文字について二種類の変換をサポートしている。

T2D エンコーディング変換

T2D エンコーディングのシンボル命令に変換する。T2D エンコーディングに基づく字母すべてを収録したフォントは、筆者の知る限り LH フォントのみである。LH フォントは教会スラヴ語文字とロシア語、ウクライナ語などの現代スラヴ語文字とでデザイン的によく相応するよう調整されている。よって学術論文などで教会スラヴ語を記述するならば T2D 変換・LH フォント使用が適していると筆者は考える。

OldSlav (G_{ik}TEX) 記法変換

OldSlav は G_{ik}TEX 教会スラヴ語フォントを pTEX でも利用できるよう筆者が調整したパッケージである。Utf82TEX は `-s` オプション指定、もしくは `<utf82tex_s>`, `</utf82tex_s>` 制御タグ指定によって G_{ik}TEX オリジナル記法によるキリル文字を OldSlav 教会スラヴ語命令に変換する。制御タグ使用時はタグ範囲内に限定して変換する。クラシックな教会スラヴ語フォントでタイプセットしたい利用者向けの機能である。

図 8 (30 頁), 図 9 (30 頁) に古スラヴ語・教会スラヴ語の直接入力例とその L^AT_EX 処理結果を示す。

```

\selectlanguage{french}
\section{フランス語 Français}
Souvent, la main portée au front, debout sur les vaisseaux, tandis que la lune se
balançait entre les mâts d'une façon irrégulière, je me suis surpris, faisant abstraction
de tout ce qui n'était pas le but que je poursuivais, m'efforçant de résoudre ce
difficile problème!

\hfill {\em Conte de Lautréamont, «Les Chants de Maldoror»}%

\hfill\today

\selectlanguage{german}
\section{ドイツ語 Deutsch}
Dann ließ sie ihre Hände langsam über meine Wangen heruntergleiten, und ihr Blick ruhte
mit unendlicher Innigkeit auf mir. Sie schüttelte den Kopf mit einem schmerzlichen
Ausdruck, als könnte sie irgend etwas nicht fassen. „Mußst du denn schon heute weg?“
fragte sie leise.

\hfill {\em A. Schnitzler, »Die Frau des Weisen«}%

\hfill\today

\selectlanguage{czech}%
\section{チエコ語 Czech}
Posláním sdružení je vytvářet předpoklady pro všestranné využívání a další rozvoj jazyka
počítačové typografie \TeX{} a příbuzného programového vybavení pro stolní tisk, zejména
mezi českými a slovenskými uživateli.

\hfill {\em Czechoslovak \TeX{} Users Group}%

\hfill\today

\selectlanguage{nippon}
\section{Latin-1, Latin-2}
\def\De#1{\bgroup\selectlanguage{german}#1\egroup}
\def\Ch#1{\bgroup\selectlanguage{czech}#1\egroup}
\De{Á}ÁÁÁÁ ÍÍÍÍ ÚÚÚÚ ĚĚĚĚ \De{Ů}ŮŮŮŮ P ÝŸÇÑ ĩ¿\
ÁÀĀĀ ĪĪ ŪŪ Ū Ū ĚĚ Ě Ě ŌŌ Ō Ō ŔŔ ŦŦ ŸŸ ŠŠ ŢŢ ĹĹ Ĺ Ĺ ŹŹ Ź Ź ĆĆ Ć Ć ŃŃ \
\De{ä}ääää ä ä ä ä \De{ü}üüü ü ü ü ü ééèè \De{ö}öööö \De{ß}ßÿÿđđčň äąää ĩi ú\Ch{ů}ů ů éě óó
řřťťýýššďďĺĺŹźžžććčň

```

図 4 ラテン文字言語入力例

フランス語 Français

Souvent, la main portée au front, debout sur les vaisseaux, tandis que la lune se balançait entre les mâts d'une façon irrégulière, je me suis surpris, faisant abstraction de tout ce qui n'était pas le but que je poursuivais, m'efforçant de résoudre ce difficile problème!

Conte de Lautréamont, «Les Chants de Maldoror»

28 juin 2009

ドイツ語 Deutsch

Dann ließ sie ihre Hände langsam über meine Wangen heruntergleiten, und ihr Blick ruhte mit unendlicher Innigkeit auf mir. Sie schüttelte den Kopf mit einem schmerzlichen Ausdruck, als könnte sie irgend etwas nicht fassen. „Mußt du denn schon heute weg?“ fragte sie leise.

A. Schnitzler, «Die Frau des Weisen»

28. Juni 2009

チェコ語 Czech

Posláním sdružení je vytvářet předpoklady pro všestranné využívání a další rozvoj jazyka počítačové typografie \TeX a příbuzného programového vybavení pro stolní tisk, zejména mezi českými a slovenskými uživateli.

Czechoslovak \TeX Users Group

28. června 2009

Latin-1, Latin-2

ÄÅÄÄÄÄ ÌÎÏ ÙÚÛÜ ÈÉÆÈÈ ÒÓÔÕÖ ð ŸĐÇŇ ĵ

ÁÀĂ Ħ Í ÚÛ Ü ÈÉÈ ÓÔ Ò ŔŖŤŸŠŚĐĎĹĹŽŽŽĆČŃŇ

äåäää ä ïî ùúü èéæèè öóôöõ ðÿýđçñ áàää ħ í úúü éè óó ŕŕŧŧŷşşđďĺĺžžžććčćň

図5 ラテン文字言語出力例

```

\selectlanguage{russian}
\section{ロシア語 %
\protect\fontencoding{OT2}\protect\selectfont{ }Русский}
\fontencoding{OT2}\selectfont%
Прежде всего откроем тайну которую Мастер не пожелал открыть Иванушке. Возлюбленную
его звали Маргаритою Николаевной. Все, что Мастер говорил о ней, было сущей правдой. Он
описал свою возлюбленную верно. Она была красива и умна.

\hfill
{\em M. Bulgakov, <Master i Margarita>}\par
\hfill\today

\selectlanguage{ukrainian}%
%%<utf82tex_r> t2a control
\section{ウクライナ語 %
\protect\fontencoding{T2A}\protect\selectfont Українська}
\fontencoding{T2A}\selectfont%
\textac{% PSCyr Akademicheskaja font
Відвідуючи меморіальні музеї Гете, Пушкіна, Толстого, кожен звертає увагу на колосальні
особисті бібліотеки цих титанів світової культури. Має і український народ подібного
велета. Це---Тарас Шевченко. }
%%</utf82tex_r>

\hfill
{\em \textlatin{\tt http://kobzar.info/our/library/}}\par
\hfill\today

\selectlanguage{nippon}%
\section{ロシア語旧正書法ほかキリル諸語}
T2A: \fontencoding{T2A}\selectfont ГгГгЗзКкКкУуУуХхЧчНнЕеӨөӨө

T2B: \fontencoding{T2B}\selectfont ЖжКкЈјЛлНнЧчЗз

T2C: \fontencoding{T2C}\selectfont ЙйБбРрБбКкЦцЧчӨөӨөМм

OT2: \fontencoding{OT2}\selectfont ЪѠV ѡv

```

図 6 キリル文字言語入力例

ロシア語 Русский

Прежде всего откроем тайну которую Мастер не пожелал открыть Иванушке. Возлюбленную его звали Маргаритою Николаевной. Все, что Мастер говорил о ней, было сущей правдой. Он описал свою возлюбленную верно. Она была красива и умна.

М. Булгаков, «Мастер и Маргарита»

28 июня 2009 г.

ウクライナ語 Українська

Відвідуючи меморіальні музеї Гете, Пушкіна, Толстого, кожен звертає увагу на колосальні особисті бібліотеки цих титанів світової культури. Має і український народ подібного велета. Це—Тарас Шевченко.

<http://kobzar.info/our/library/>

28 червня 2009 р.

ロシア語旧正書法ほかキリル諸語

T2A: Г г F f З з K k Y y X x Ч ч H h E e Э э O o

T2B: Ж ж Ы ы Л л Ъ ъ Н н Ч ч З з

T2C: Й й Б б Р р Б б К к Ц ц Ч ч Е е Ё ё М м

OT2: Ё ё V v

図7 キリル文字言語出力例

古典ギリシア語

Ἄνδρα μοι ἔννεπε, Μοῦσα, πολύτροπον, ὃς μάλα πολλά
 πλάγχθη, ἐπεὶ Τροίης ἱερὸν πτολίεθρον ἔπερσε.
 πολλῶν δ' ἀνθρώπων ἴδεν ἄστεα καὶ νόον ἔγνω,
 πολλὰ δ' ὃ γ' ἐν πόντῳ πάθεν ἄλγεα ὃν κατὰ θυμόν,
 ἀρνύμενος ἣν τε ψυχήν καὶ νόστον ἐταίρων.
 ἀλλ' οὐδ' ὣς ἐτάρους ἐρρύσατο, ἰέμενός περ·
 αὐτῶν γὰρ σφετέρῃσιν ἀτασθαλίῃσιν ὄλοντο,
 νήπιοι, οἳ κατὰ βοῦς Ἵπερίονος Ἥελίοιο
 ἤσθιον· αὐτὰρ ὁ τοῖσιν ἀφείλετο νόστιμον ἦμαρ.
 τῶν ἀμόθεν γε, θεά, θύγατερ Διός, εἰπέ καὶ ἡμῖν.

[Ὅμηρος]

ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ αβγδεζηθικλμνξοπρσ τυφχψω
 ΆΆΆΆΆΆΆ άάάάάάά ΈΈΈΈΈ Έέέέέ ΉΉΉΉΉΉΉ ήήήήήήήήήή
 ΊΊΊΊΊΊΊ ιίίίίίίίί ΌΌΌΌΌΌ όόόόόό ό ΎΎΎΎΎ ύύύύύύύ
 ΩΩΩΩΩΩΩ ώώώώώώώ άά έέ ήή ίί όό ύύ ώώ
 Ά₁Ά₁Ά₁Ά₁Ά₁Ά₁Ά₁ άάάάάάάάά Ή₁Ή₁Ή₁Ή₁Ή₁Ή₁Ή₁ ήήήήήήήήήή
 Ό₁Ό₁Ό₁Ό₁Ό₁Ό₁Ό₁ όόόόόόόόό ΆΆΆΆΆ άάάάάάά
 ΈΈ ΉΉΉ ήήήήή ΊΊΊ ιίίίίίί
 ύύ ύύ ρρ ύύ ΎΎ ΎΎ Ρ Ρ ΌΌΌΌΌ₁ όόόόόό
 ΆΈΉΊΌΎ ΊΎ άέήίόύώ ύ ύ β θ ϒ ϒ ϒ φωχ Ϙ ϒ ϒ F F Ϙ λ λ κ ρ ς j ε ε δ'
 Greek Accents
 ´ ˘ ˙ ˚ ˛ ˜ ˝ ˞ ˟ ˠ ˡ ˢ ˣ ˤ ˥ ˦ ˧ ˨ ˩ ˪ ˫ ˬ ˭ ˮ ˯ ˰ ˱ ˲ ˳ ˴ ˵ ˶ ˷ ˸ ˹ ˺ ˻ ˼ ˽ ˾ ˿ ˿

28 Ιουνίου 2009

図 11 ギリシア語出力例

7.6 タイ語

タイ語には単語の境界でのみ改行が許容されるという正書法がある。-t もしくは -T オプションを指定すると、`Utf82TeX` は、Unicode のタイ文字からなるテキスト断片を TIS-620 タイ文字集合に変換し、タイ語単語境界マーキングソフトウェアである `swath` を呼び出して単語境界にブレイクポイントを挿入する。UNIX の場合標準では `{\wbr}` コントロールシーケンスが挿入される。このマーキング部分で必要に応じて改行されるようにするためには、`{\wbr}` に対してこれを実現する定義をしておく必要がある。図 12 (33 頁) がその定義である。これは単にグルーを挿入しているだけである。例ではグルーの延び縮みは 0.6pt に設定している。“Overfull \hbox” メッセージが出て、行端がはみ出してしまうような場合、この数値を大きくして調整するとよい。

```
\def\wbr{\hskip0pt plus0.6pt minus0.6pt\relax}%
```

図 12 タイ語単語境界分割定義

タイ語 `ThaiLATEX` を前提としている^{*19}。ThaiL^AT_EX には Babel タイ言語定義 `thai.ldf` が添付されている。`\today` タイ歴の年月日、キャプションを出力したい場合は、図 13 (33 頁) のように `thai.ldf` を トノ氏作成 `8bt2ptex.pl` によって「^^十六進数」形式に変換しておくことタイ語が文字化けしない^{*20}。

```
# cd /usr/local/teTeX/share/texmf-local/tex/generic/babel
# cp -p thai.ldf thai.ldf.orig
# 8bt2ptex.pl thai.ldf.orig > thai.ldf
```

図 13 thai.ldf 変換

図 14 (34 頁)、図 15 (34 頁) にタイ語の直接入力例とその L^AT_EX 処理結果を示す。

^{*19} <http://linux.thai.net/>. ダウンロードは <ftp://linux.thai.net/pub/thailinux/software/thailatex> から UNIX 用最新版が入手できる。Windows 版は <http://miktexthai.sourceforge.net/> から取得する。

^{*20} `8bt2ptex.pl` スクリプトは稲垣徹氏のサイト <http://www2.tba.t-com.ne.jp/ing/> からダウンロード可能である。

```
\selectlanguage{thai}%  
\section{タイ語 ภาษาไทย}  
\def\wbr{\hskip0pt plus0.6pt minus0.6pt\relax}%  
\fontencoding{LTH}\fontfamily{norasi}\selectfont%  
เป็นมนุษย์สุดประเสริฐเลิศคุณค่ากว่าบรรดาฝูงสัตว์เดรัจฉานจงฝ่าฟันพัฒนาวิชาการอย่าล้างผลาญญาเช่นฆ่าบีทา  
ใครไม่ถึงโทษโกรธแข่งชดฮึดฮัดด่าหัดดอภัยเหมือนกีฬาอัชฌาสัยปฏิบัติประพฤติกฏกำหนดใจพูดจาให้จ๊ะจ๋า จ่าๆ  
น่าฟังเอ๋ย  
\hfill\today
```

图 14 タイ語入力例

タイ語 ภาษาไทย

เป็นมนุษย์สุดประเสริฐเลิศคุณค่ากว่าบรรดาฝูงสัตว์เดรัจฉานจงฝ่าฟันพัฒนาวิชาการอย่าล้างผลาญญาเช่นฆ่าบีทาใคร
ไม่ถึงโทษโกรธแข่งชดฮึดฮัดด่าหัดดอภัยเหมือนกีฬาอัชฌาสัยปฏิบัติประพฤติกฏกำหนดใจพูดจาให้จ๊ะจ๋า จ่าๆ น่าฟัง
เอ๋ย

28 มิถุนายน พ.ศ. 2552

图 15 タイ語出力例

7.7 ヴェトナム語

ヴェトナム語の変換規則は拡張変換テーブルで実現しており、**-x オプション**付きで実行する。
 図 16 (35 頁), 図 17 (35 頁) にヴェトナム語の直接入力例とその L^AT_EX 処理結果を示す^{*21}。

```
\selectlanguage{vietnam}%
\section{ヴェトナム語 %
{\protect\fontencoding{T5}\protect\selectfont Tiếng Việt}}
\fontencoding{T5}\selectfont%
Phần “Những câu hỏi và giải đáp thường gặp”(viết tắt từ tiếng Anh là FAQ) được nêu ra ở
đây nhằm mục đích thu thập những câu hỏi thường gặp trong thực tế và những lời giải đáp
thích hợp nhất của nó. Từ lần ấn bản cuối cùng đến nay, đã có những sự thay đổi khá lớn
trong cấu trúc của FAQ. \textbf{Để hiểu rõ hơn bạn nên xem lại chương “Làm sao tôi có thể
đọc được FAQ”.}

\hfill{\em CJK samples}}\par
\hfill\today
```

図 16 ヴェトナム語入力例

ヴェトナム語 Tiếng Việt

Phần “Những câu hỏi và giải đáp thường gặp” (viết tắt từ tiếng Anh là FAQ) được nêu ra ở đây nhằm mục đích thu thập những câu hỏi thường gặp trong thực tế và những lời giải đáp thích hợp nhất của nó. Từ lần ấn bản cuối cùng đến nay, đã có những sự thay đổi khá lớn trong cấu trúc của FAQ. **Để hiểu rõ hơn bạn nên xem lại chương “Làm sao tôi có thể đọc được FAQ”.**

CJK samples

Ngày 28 tháng sáu năm 2009

図 17 ヴェトナム語出力例

^{*21} Vn_TE_X-3.02 添付のフォント定義 `t5cmtt.fd` にはバグがある。10 行目 “<5><6><7><8>vntt8<9>cmtt9%” の `cmtt9` は誤りであって、`vntt9` が正しい。直接入力例では修正したもので出力した。

7.8 Unicode 漢字, 異体字, 機種依存文字

「鷗」や「禱」などの JIS 第一・二水準で定義されていない Unicode 漢字を `\UTFM{9dd7}` や `\UTFM{79b1}` に変換する。「▼」を前置することで異体字を出力できる場合がある。

記号類, 半角カナなど機種依存文字として知られる文字の変換が可能である。これは大友康寛氏が提供くださった拡張変換テーブルで実現しており, `-x オプション` 付きで実行する。

プリアンブルに OTF パッケージを指定しておく。

図 18 (36 頁), 図 19 (36 頁) に直接入力例とその L^AT_EX 処理結果を示す。

森鷗外, 祈禱書, ▼吉野家, 内田百▼間, ▼高島屋, 高島屋

①②ⓂⅧⅧ 鷗禱書kmmgTeL (株)(有)(代) 鷗禱書禱禱 𐄂 ≡ ∫ ∫ 續襲鎡銚𐄂𐄂 アイエオカクコ

図 18 Unicode 漢字・異体字・機種依存文字入力例

森鷗外, 祈禱書, 吉野家, 内田百間, 高島屋, 高島屋

①②ⓂⅧⅧ 鷗禱書kmmgTeL (株)(有)(代) 鷗禱書禱禱 𐄂 ≡ ∫ ∫ 續襲鎡銚𐄂𐄂 アイエオカクコ

図 19 Unicode 漢字・異体字・機種依存文字出力例

7.9 中韓文字

OTF パッケージは `\UTFC`, `\UTFE`, `\UTFK` コントロールシーケンスによってそれぞれ中国語簡体字, 中国語繁体字, 韓国語の漢字を優先的に拾う機能を有している。`-c C|T|K オプション` または `<utf82tex_c|t|k>` 制御タグ (| の部分は中国語簡体字, 中国語繁体字, 韓国語に応じて選択) によって漢字を `\UTFx` シーケンスに変換する。句読点 (.,.,.) についてのみ `\penalty10000` を前置してそこで改行されないよう配慮している。

ハンゲルと判断できる文字は標準で `\UTFK` コントロールシーケンスに変換する。

プリアンブルに OTF パッケージを指定しておく。

図 20 (37 頁), 図 21 (38 頁) に直接入力例とその L^AT_EX 処理結果を示す。中韓の例文はいずれも CJK パッケージのサンプルから採ったものである。

```
\section{日本語}
雪後庵は起伏の多い小石川界隈の高臺にあつて、幸ひに戦災を免かれた。
三千坪に及ぶ名高い小堀遠洲流の名園と共に、京都のとある名刹から移された中雀門も、奈良の古い寺をそのま
>移した玄関や客殿も、あとに建てられた大廣間も、何一つ損なはれてゐなかつた。
```

戦後の財産税さわぎの只中に、雪後庵は元の持主の實業家の茶人の手から、美しい元氣な女主人の手に渡つて、たちまち名高い料理屋になつた。

```
\vspace{1zh}%
\hfill 三島由紀夫『宴のあと』
```

```
\vspace{1zh}%
骨練平。
```

```
%%<utf82tex_k>
\section{韓國語}
이 FAQ 은 자주 반복되는 질문과 그에 대한 대답을 간단명료한 양식으로 모아 엮어졌습니다.
이 FAQ 의 구조는 지난 판에 비하여 획기적으로 변경되었습니다.
상세한 것은 “이 FAQ 을 어떻게 읽을 것인가” 라는 대목을 참조하시기 바랍니다.
```

```
\vspace{1zh}%
骨練平。
%%</utf82tex>
```

```
%%<utf82tex_c>
\section{简体中文}
本常問問答集 (FAQ list)
是从一些經常被問到的問題及其適當的解答中、以方便的形式摘要而出的。
跟上一版不同的是、其編排結構已徹底改變。
有關新結構的細節、可參考如何閱讀本問答集及了解其編排結構該項中的說明。
```

```
\vspace{1zh}%
骨練平。
%%</utf82tex>
```

```
%%<utf82tex_t>
\section{繁體中文}
本常問問答集 (FAQ list)
是從一些經常被問到的問題及其適當的解答中、以方便的形式摘要而出的。
跟上一版不同的是、其編排結構已徹底改變。
有關新結構的細節、可參考如何閱讀本問答集及了解其編排結構該項中的說明。
```

```
\vspace{1zh}%
骨練平。
%%</utf82tex>
```

図 20 CJK 統合漢字入力例

日本語

雪後庵は起伏の多い小石川界限の高臺にあつて、幸ひに戦災を免かれた。三千坪に及ぶ名高い小堀遠洲流の名園と共に、京都のとある名利から移された中雀門も、奈良の古い寺をそのまま移した玄関や客殿も、あとに建てられた大廣間も、何一つ損なはれてゐなかつた。

戦後の財産税さわぎの只中に、雪後庵は元の持主の實業家の茶人の手から、美しい元氣な女主人の手に渡つて、たちまち名高い料理屋になつた。

三島由紀夫『宴のあと』

骨練平。

韓國語

이 FAQ 은 자주 반복되는 질문과 그에 대한 대답을 간단명료한 양식으로 모아 엮어졌습니다 . 이 FAQ 의 구조는 지난 판에 비하여 획기적으로 변경되었습니다 . 상세한 것은 “이 FAQ 을 어떻게 읽을 것인가” 라는 대목을 참조하시기 바랍니다 .

骨練平。

简体中文

本常問問答集 (FAQ list) 是从一些经常被问到的问题及其适当的解答中、以方便的形式摘要而出的。跟上一版不同的是、其编排结构已彻底改变。有关新结构的细节、可参考如何阅读本问答集及了解其编排结构该项中的说明。

骨练平。

繁體中文

本常問問答集 (FAQ list) 是從一些經常被問到的問題及其適當的解答中、以方便的形式摘要而出的。跟上一版不同的是、其編排結構已徹底改變。有關新結構的細節、可參考如何閱讀本問答集及了解其編排結構該項中的說明。

骨練平。

図 21 CJK 統合漢字出力例

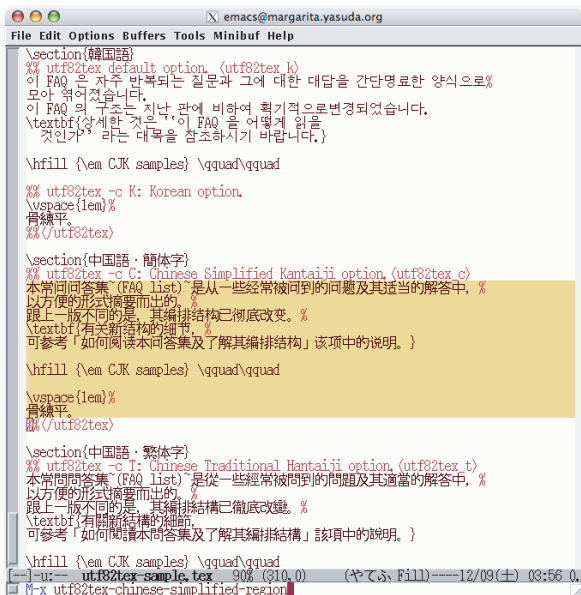


図 22 Emacs 操作 (変換前)

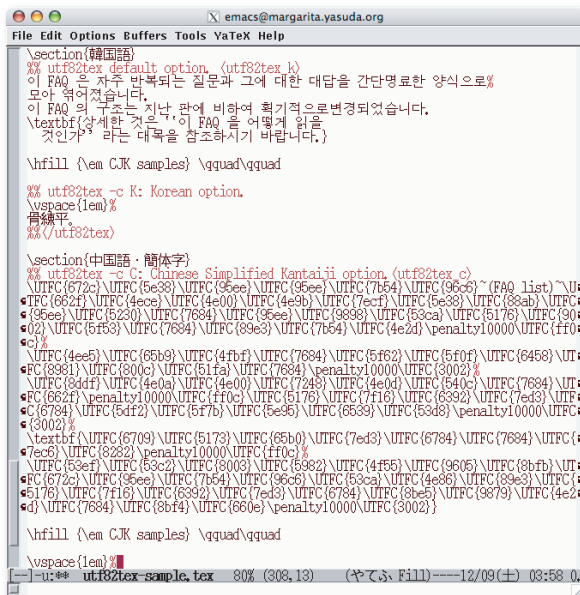


図 23 Emacs 操作 (変換後)

7.10 Emacs 環境

GNU Emacs での変換は 1_TE_X 文書ソースを作成しつつ、リージョンに対し適宜必要なオプションを指定して `Utf82TEX` を実行するのに向いている. `-c オプション` で `\UTFK`, `\UTFC` などに CJK 漢字を変換したい場合, コマンドラインからファイルに対し実行するとテキストすべてに適用されて, 通常の日語も `\UTFK` のシーケンスに変わってしまう. こういうときは Emacs でリージョンを指定して必要な部分のみを変換するのがよい. もちろん, `<utf82tex_x>` タグで制御してもよい.

Emacs 環境での変換前後の様子をそれぞれ図 22 (39 頁) 及び図 23 (39 頁) に示す.

8 ユーティリティ

8.1 ulatex

Utf82TeX の出力は UTF-8 エンコードであり、さらに EUC などの日本語文字コードに変換して pL^AT_EX 処理を行う必要がある (40 頁, 図 24).

```
% utf82tex -tx < src.tex | iconv -c -f UTF-8 -t ISO-2022-JP > src.jis
% platex src.jis
% dvipdfmx -l src.dvi
```

図 24 Utf82TeX コマンドライン実行例

原稿修正のたびに図 24 (40 頁) のようにいちいち `utf82tex`, `iconv`, `platex`, `dvipdfmx` コマンド群をタイプするのは億劫だという利用者は, `ulatex` ユーティリティを利用できる. `ulatex <UTF-8 TEX ファイル>` を実行することによって, 同一ジョブ名の PDF 生成まで自動化できる. ファイルの `.tex` 拡張子は省略できる. `ulatex` は UTF-8 の T_EX ファイルを処理する簡易 `platex` コマンドのようなものである.

`ulatex` ユーティリティは UNIX 用シェルスクリプト, Windows 用バッチファイルとして添付されている. 単純なものなので, PS 変換を加える等, 利用者の好みでカスタマイズしていただきたい. 変数の設定でパス値をインストール環境に合わせておく必要がある.

UNIX 版 `ulatex` は引数が複数指定されると, 最後のものは T_EX ファイルとして扱い, 残りを `dvipdfmx` に渡すようになっている. `ulatex -l -p b5 src` とすると “-l -p b5” オプション付きで `dvipdfmx` が起動され, landscape B5 版サイズで PDF `src.pdf` が生成される.

UNIX 版 `ulatex` では, コンパイル・ログに `references warning` が出力されている場合, 未解決参照があるものとしてこれが解決するかまたは最大 `MAXC` 回数 (5 回) まで `platex` を再実行するようになっている.

8.2 genchrtbl

添付 `Makefile`, `install.bat` で拡張テーブル生成支援ユーティリティ `utils/genchrtbl` がインストールされる. これは拡張テーブルのテンプレートを生成するツールである. 作成したい Unicode 文字のコードポイント範囲を十六進数で指定すると, その指定コードポイント範囲で次のような形式でテーブルの雛形テキストを生成し, 標準出力に書き出す.

```
e UTF-8 文字 * # U+ コードポイント Unicode 文字名称
```


出力されたテキストを Unicode 対応エディタで開き，“*”の部分に対してユーザの変換したい T_EX シーケンスを記述していけばよい。Perl Unicode 実装においてコードポイントの文字が未定義の場合、テキスト行頭に # を付加し、コメント行として生成する。# 以降の文字列 (Unicode コードポイント, 文字名称) はコメントであり、削除・修正しても構わない。

例えば、以下のように実行すると、キリル文字のテーブルテンプレート `cyrtempl.tbl` が生成される。

```
% genchrtbl 0400 04ff > cyrtempl.tbl
% cat cyrtempl.tbl
e È * # U+0400: CYRILLIC CAPITAL LETTER IE WITH GRAVE
e Ë * # U+0401: CYRILLIC CAPITAL LETTER IO
e Ъ * # U+0402: CYRILLIC CAPITAL LETTER DJE
e Ѓ * # U+0403: CYRILLIC CAPITAL LETTER GJE
e Є * # U+0404: CYRILLIC CAPITAL LETTER UKRAINIAN IE
.....
```

9 その他

9.1 利用条件

本ソフトウェアはフリーソフトウェアで無償にて利用できる。GNU GPL を適用する。本ソフトウェアに含まれるプログラム、ドキュメントの運用結果に関し作者及びその関係者はいかなる責任も負わない。利用者の責任でご活用いただきたい。

Utf82T_EX はベータ版である。「巧遅拙速に如かず」、「今日という日を摘め」との言を頼りに乏しい知識をも省みず拵えたところがあり、文字とその変換結果の対応関係において自信のない部分もある。ご指摘をお待ちする。ご意見、ご質問は題記メールアドレスまでご連絡いただけると幸いである。

9.2 謝辞

パッケージに同梱されているグルジア語／ヴェトナム語／機種依存文字変換用拡張テーブルは、大友康寛氏より提供された Perl ソースコードから作成したものである。

CJK Unicode テーブルは角藤亮氏 `utf8toutf` のコードを参考にさせていただいた。

大友氏、角藤氏に深謝申し上げたい。

9.3 misima

筆者は、T_EX プリプロセッサとして、`misima` 旧仮名遣い・旧字変換支援を別に公開している。`misima` は和文表記変換を主機能とするが、Utf82T_EX と同等の T_EX 多言語文字変換機能を実装している上、さらに古典籍組版向けの変換機能 (下記参照) を有している。

`misima` は Perl 5.8.0 以上、日本語形態素解析システム『茶釜』2.3.3^{*22} がインストールされた環境で動作す

^{*22} <http://chasen.naist.jp/hiki/ChaSen/>

る。利用者の計算機に misima をインストールしなくとも、ブラウザ、GNU Emacs、コマンドラインから利用できる Web サービスを作者のサイトで公開している。現代表記向けに調整された IME を使いつつ、旧仮名遣い、旧字体の文章を大量に引用しなければならない TeX 利用者の課題・利便性を考慮したつもりである^{*23}。

- 新字、略字を旧字体、異体字に変換する。Unicode にない異体字は OTF パッケージ `\CID` 命令参照で出力することができる。
- 一定のシーケンスからなるベタの文字列を旧字体に変換し、藤田眞作氏の漢文訓点パッケージ `sfkanbun`^{*24} のフォーマットに整形する。入力と出力をそれぞれ図 25 (42 頁) 及び図 26 (43 頁) に示す。

```

\DeclareAJKundokuStyle%
\let\ninojiten=\ajvarNinoji%

<misima_kanbun>
顔淵・季路待ス。子曰ハク、盍なんゾ3ルト各々言ハ2爾なんヂノ志ヲ1。
</misima_kanbun>

\DeclareOriginalKundokuStyle%
<misima_kanbun>
楽シム G 得テ2天下之英オヲ1而教 C-育スルヲ之ヲ J。
</misima_kanbun>

<misima_kanbun>
聞ク H 吏部ノ有リ G 以テ2博学宏くわう辞ジヲ1選ブ者 J, 人尤もつとモ謂いヒ2之ヲオト1, 得ウト O 美仕ヲ
K。
</misima_kanbun>

<misima_kanbun>
既ニ出デテ得2其船ヲ1, 便チ扶そヒ2向さきノ路みちニ1, 処処ニ誌しス R 之ヲ。
及ビ2郡下ニ1, 詣いたリテ2太守ニ1, 説クコト如シ R 此かくノ。
太守即すなはチ遣しムルモ G 人ヲシテ随したがヒテ2其ノ往クニ1, 尋ネ C 向ニ所ヲ JR 誌セシ, 遂ニ迷ヒテ不
2復得1 R 路ヲ。
</misima_kanbun>

<misima_kanbun>
闔かふ廬りよ曰ハク、「可キ3試ミニ以テス2婦人ヲ1乎」。
曰ハク、「可ナリ」。
於テ R 是ここニ許ス R 之ヲ。
</misima_kanbun>

```

図 25 漢文訓点変換 (入力)

- 和文繰返し表現を OTF パッケージ「くノ字点」`\ajKunoji`, `\ajDKunoji` に変換する。例は図 27 (43 頁) を参照。

^{*23} 詳細は <http://yasuda.homeip.net/misima/> を参照。

^{*24} <http://imt.chem.kit.ac.jp/fujita/fujitas2/texlatex/tategumi/sfkanbun.sty>

於_レ是_ニ許_レ之_ヲ。
 闔_カ廬_{ハク}曰_ク、「可_キ試_ミ以_テ婦_ノ人_ニ乎_ヲ」。曰_ク、「可_{ナリ}」。
 既_ニ出_デ得_テ其_ノ船_ヲ、便_チ扶_ヒ向_キ路_ヲ處_ニ處_ニ誌_レ之_ヲ。及_ビ
 郡_ノ下_ニ詣_リ太_ノ守_ヲ、說_ク如_ク此_ノ。太_ノ守_即遣_ハ
 人_ヲ隨_ヒ其_ノ往_キ、尋_ヒ向_キ所_ヲ誌_シ遂_ニ迷_ヒ不_レ復_ス
 得_レ路_ヲ。
 才_ト得_テ美_シ仕_メ甲_ヲ。
 聞_ク吏_ノ部_有以_テ博_ク學_宏辭_ヲ選_ブ者_上、人_尤謂_フ之_ヲ
 樂_シ得_テ天_ノ下_之英_才而_テ教_育之_ヲ上_。
 顏_淵・季_路待_ス子_曰、盍_ニ各_々言_フ爾_ノ志_ヲ。
 樂_シ得_テ天_ノ下_之英_才而_テ教_育之_ヲ上_。

図 26 漢文訓点変換 (出力)

そろそろ
 それぞれ
 散り散り
 そろく
 それく
 散りく

図 27 くの子点変換

9.4 変更履歴

- 2003/05/03 初版公開
- 2004/04/25 スラヴ語, ギリシア語サポート
- 2005/12/28 0512 タイ語, 拡張テーブル, 中韓制御タグサポート
- 2006/06/09 0606 教会スラヴ語, Windows サポート, ulatex, genchrtbl ユーティリティ
- 2006/12/10 0612 中韓 \penalty10000 挿入
- 2007/01/26 0701 BibTeX アクセント記法を考慮したテーブル仕様変更
- 2007/07/14 0707 古典ギリシア語例中の省音 (elision) を訂正
- 2007/11/12 0711 ドイツ語・BibTeX オプション, 秀丸・Jedit X・TeXShop インタフェース
- 2009/06/28 0906 Unicode CJK Extension B 漢字サポート

参考文献

- [1] トニー・グラハム著 乾和志, 海老塚徹訳 『Unicode 標準入門』翔泳社, 2001.
- [2] Jenkins J. H., Cook R., 井作恆, 曲理查, “UNICODE HAN DATABASE (UNIHAN)“, <http://www.unicode.org/reports/tr38/tr38-5.html>, 28 Mar. 2008.
- [3] <http://www.unicode.org/>, Unihan database, Unicode version 5.1.0, 3 Mar. 2008.
- [4] Adobe Systems Incorporated, “Adobe CMap and CIDFont Files Specification”, Version 1.0, 11 June 1993.

以上

Copyright © 2006–2009, isao yasuda, All Rights Reserved.