

中期インド・アリアン古文献解析ツールの リナックスへの拡張

逢 坂 雄 美

我々は情報科学研究者と言語学研究者との共同研究の下に、中期インド・アリアン古文献を系統的に解析できるツールを、始めにマッキントッシュ上で、次いでウィンドウズ上で開発した。解析システムはフォント、エディタ、解析ツール（テキストデータ、韻律解析、索引作成）の3部構成になっている。この計算機リソースは、独自の特殊フォント系（=Pali96:通常のローマ字フォント+拡張アスキーコード上の特殊フォントからなる）をもとにテキストデータを作成し、言語解析において重要な語彙索引・詩脚索引・韻律を解析できる。本論文では、今後その使用頻度が高まると予測される LINUX 上でのツール開発について議論する。

Computer Analysis Tool in Linux for Manuscripts in Middle Indo-Aryan

Yumi OUSAKA

In corroboration with the linguistic scholars and computer scientists, we have made the computer tools for the systematic analysis of the texts in Middle Indo-Aryan on Macintosh and then on Windows PC. Our system consists of three parts: the special font system, the text editor, and analysis tools for metres and the tools for the productions of the word index and the pada-index. This system works well on not only Macintosh but Windows. In this paper we will discuss on the development of the analysis system on Linux, which would be available more and more in future.

仙台電波工業高等専門学校
〒989-3128 仙台市青葉区愛子中央4-16-1
Sendai National College of Technology
Ayashi Chuo 4-16-1, Aobaku, Sendai 989-3128

1. 序論

我々は情報科学研究者と言語学研究者との共同研究の下に、中期インド・アリアン古文献を系統的に解析できるツールを、始めにマッキントッシュ上で、次いでウィンドウズ上で開発した[1]. 図1に解析システムの全体構成を示す. 我々の解析システムはフォント, エディタ, 解析ツール(テキストデータ, 韻律解析, 索引作成)の3部構成になっている. この計算機リソースは, 独自の特殊フォント系 (=Pali96:通常のローマ字フォント+拡張アスキーコード上の特殊フォントからなる)をもとにテキストデータを作成し, 言語解析において重要な語彙索引・詩脚索引・韻律を解析できる. 韻律解析結果を使用することにより, 各詩偈を効率よく詩脚に分解でき, この分割を利用することにより詩脚索引を作成できる. これらのツールを使用してこれまで, 20冊以上の基礎資料[2]を出版すると同時に, 作成した電子化テキスト・電子ブック等のデータをWEB上に公開した[www.sendai-ct.ac.jp/~ousaka].

これまでの解析ツールは情報学研究者が操作していたが, 韻律解析・索引作成プログラム等をワープロ並にコンパクトな使いやすい形にすると, 言語学者自身が直接使用できると考えられる. この結果, 必要な第1次基礎資料を必要な時に, 言語学者が自身のパソコン上に手軽に作成・表示できることになり, 一層研究が進展すると予想される. この目的で, 最近, 従来のツールを改良して, 言語学者が情報学研究者の助けを借りずに, 自身で自由に使える計算機処理システムをマッキントッシュとウィンドウズ上に再構築した[3]. これらの解析ツールをCDに収録し(又同一のデータを上記サイトにて)公開している.

これらのツールは, 従来のClassic Mac 又はWindows98系統のOS上では支障なく動作する. 本論文では, 今後その使用頻度が高まる

と予測されるLinux上でのツール開発について議論する.

第2節では, フォント系と使用するLinux Distribution について記述する. 第3節ではLinux のエディタについて述べ, 第4節では解析ツールの移植について議論する. 第5節では問題点等について議論する.

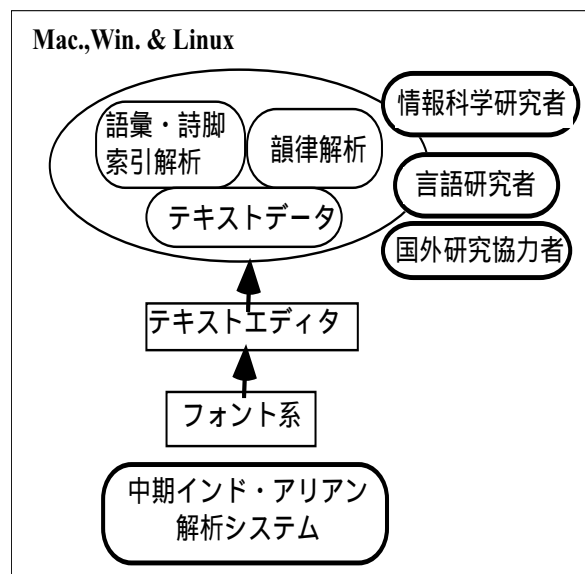


図1. 中期インド・アリアン古文献の解析システム

2. フォント系と使用するLinux Distribution

我々の解析で使用しているフォント系 Pali96 は元々, マッキントッシュでTrue Type フォントとして作成され, その後Windowsに移植された. 両OSにおいて, このフォント系の各文字のアスキーコード割り当ては同じであり, アスキーコードで128~255番の領域(所謂, 拡張アスキーコード, Linuxでは非アスキー文字)まで定義されている特殊なフォント系である (Windowsではsymbol fontとして扱われている). 表1はWindowsにおける, 我々の作成したフォントのアスキーコードへの割り当てと共に, これらのフォントのキー操作法を示している. WindowsPCの場合マッキントッシュと違って, 特殊フォントの簡便なキーボード入力法がないので, 直

接このフォントを使ってテキストを作成するのは非常に能率が悪い。この為の簡便なエディタ EDISAN を開発した。EDISAN により、マッキントッシュのテキスト入力と同じ様な簡単さでできる。なお、マッキントッシュでは、殆どのワープロソフトで簡単にこのフォントのキー入力ができる。

図1の解析システム構築の要点は、同一のキーコード割り当てを持つフォント系の使用にある。最近のLinux Distribution では、WindowsのTrue Type フォントを使用できるようになった。True Type フォントの使用方法については、xfs (フォントサーバ) を使用するLinuxとそれ以外のものがある。フォントサーバを使用するDistribution 中からRed Hat Linux 7.2を採用した。この理由として、フォントインストール及びその設定が比較的簡単にかつ正確にできること[4]、解析ツールの開発ツールKylix (Windows DelphiのLinux版)が正常に動作すること(第4節参照)、Red Hat Linux 7.2を簡単に入手できる

こと、が挙げられる。

次節でインストールされたフォントの使用方法について議論するが、上記の表1に述べたようなキー入力できる簡易エディタがLinuxには現在のところ無いようである。

3. Editor, OpenOffice.org 1.0

Red Hat Linux 7.2上の使用可能なエディタについては有名な Emacs (Mule) や OpenOffice.org 1.0 (正確にはそのOffice ツールの中のWriter) 等について検討した (WindowsPCで開発した簡易エディタEDISANを拡張することについては第5節参照)。通常のワープロ感覚で使用できるエディタを採用するという観点から、我々はフリーソフトウェアOpenOffice.orgを使用することにした。このツールはMac, Windows, Linuxでともに動作し、マイクロソフト社のOfficeツールと互換性があり、非常に優れている。ちなみに、図2と3はScreen Shotを取り、その結果をOpenOffice.org経由でマッキントッシュに取

表1. WindowsPCにおけるパーリフォントPali96のキー操作法とアスキーコード割り当て。
 正規の文字 a, i 等は、勿論通常通りにタイプできる。特殊文字 “ã” はAlt keyと正規の文字 “a”を同時に押すことによりタイプできる。同様に他の特殊文字 “u”, “kh” 等を打つことができる。特殊文字 “ĩ” は Alt keyと正規の文字 “i”を同時に押した後で引続き文字 “i”を打つことによりタイプできる。同様に “ũ” を打つことができる。最後に、特殊文字 “th” は3つのキー、Alt key, Shift key, 正規の文字 “y”を同時に押すことによりタイプできる。

character	key stroke	numeric code	character	key stroke	numeric code
ã	<A>+n, a	136	ţ	<A>+t	160
ā	<A>+a	140	th	<A>+<S>+y	231
ĩ	<A>+` , i	147	đ	<A>+d	182
ī	<A>+i, i	148	đh	<A>+i, U	243
ũ	<A>+e, u	156	ñ	<A>+n, n	150
ū	<A>+u, u	159	th	<A>+y	180
ř	<A>+r	168	dh	<A>+z	189
ṃ	<A>+m	181	ph	<A>+p	185
kh	<A>+k	251	bh	<A>+b	186
gh	<A>+9	187	l̇	<A>+l	194
ñ	<A>+g	169	l̇	<A>+<S>+l	241
ch	<A>+c	141	ś	<A>+s	167
jh	<A>+j	198	ş	<A>+x	197
ñ̇	<A>+<S>+i	246	ḣ	<A>+h	250

り込んでいます。

図2にOpenOffice.orgの概形を示している。この図のテキスト部分は、OpenOffice.orgを使用して入力された特殊フォント部分も含んだ結果を示している。InsertメニューのSpecial Characters.....をクリックすることにより、文字割り当てテーブル(図2の中の表)が現れるので、必要な文字をクリックすることにより、特殊フォントをテキストフィールド(この表の下)のテキスト群)に入力できる。今のところ残念ながら、特殊フォントに対して表1のような簡単なキー入力法が存在しない(このことについては今後検討する予定である)。しかし

ながら、一応特殊フォントを正確に入力でき、得られたテキストファイルは次節で議論する索引作成ツールの入力としても利用できる。又その結果得られる特殊フォントを含む索引の計算結果を正確に表現できる。これらのことから、Red Hat Linux 7.2では、OpenOffice.orgが我々の解析システムで必要とするエディタとしての機能がある程度満たしていると言える。

我々の解析ツールを使って得られた結果がセーブされたファイルをLinuxで正しくオープンすることは、以下のようにしてできる。最初に、OpenOffice.orgを立ち上げておき、次にファイルをオープンする。そうすると通

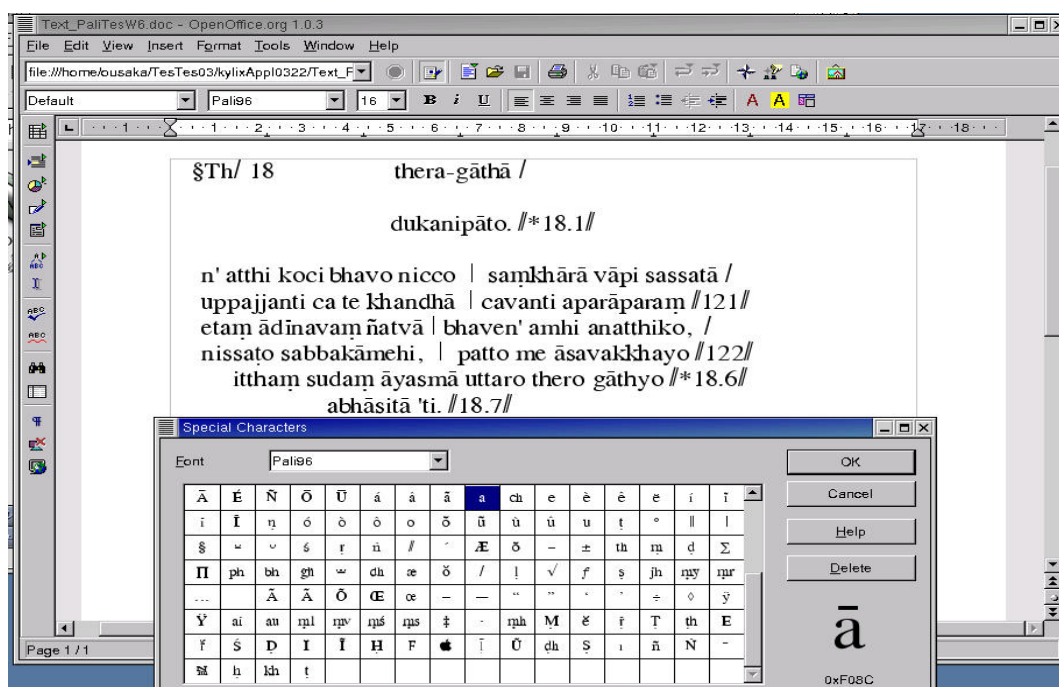


図2. OpenOffice.org1.0の概形

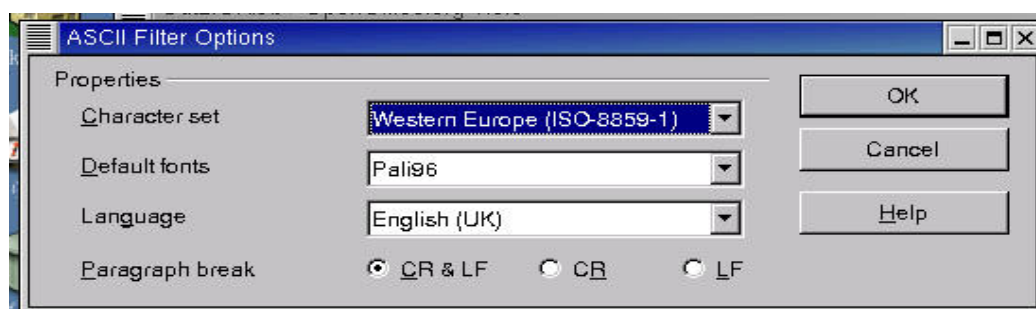


図3. ASCII Filter Option

常のダイアログボックスがオープンされるので、適切なファイル名を入れてOpenボタンをクリックする。そうすると、次に自動的に図3のASCII Filter Option ボックスがオープンするので、図のように各項目を選択し、OKボタンをクリックする。この結果、特殊フォントを正確に表示した形でファイルを開くことができる。なお、Character setの項目には、Western Europe(Iso-8859-1)を指定しておく必要がある。図3のASCII Filter Option ボックスがオープンされないときは、ファイルオープンのダイアログボックスのFile Types の項目で、Text Encoded (OpenOffice.org Writer/Web)を選択する必要がある。

上記の方法は、特殊フォントを使用して作成されたWindows ファイルをオープンするときにも活用できる。

4. 解析ツールの移植

Windows PC では解析ツールの作成アプリケーションとしては、Delphi (Borland)を使用している。Red Hat Linux 7.2では、アプリケーションKylix3.0 (Borland) が使用可能であり、これはDelphi とよく対応している。Delphi とKylix3.0は殆ど同一の機能を持ち、Windows の解析ツールの移植は、Windows OSの独自の機能をを使用していない限り、次のような変更により簡単にできる。(1) Implementation部で、{\$R *.dfm}から{\$R *.xfm}に変更する。(2) すべてのuses節において、使用しているDelphi用CLXユニットを対応するKylix用VCLユニットに変更する。(3)各プログラムユニットのパーミッションの解除に注意して、Kylixでプロジェクトファイルを開く

て、ツールを実行すればよい。

これまで詩脚索引ツールに対して、この移植作業を実行している。図4に、このツールを実行したときの、ファイル入出力ダイアログを示している。この画面は、Windows PCでの結果と完全に対応しており、この入出力ダイアログは正常に動作することを確認している。前節で述べたように、特殊フォントを使用して表されている計算結果は、OpenOffice.org を使うことにより文字化け等が無く正常に表示可能である。この結果、我々の解析ツールで重要な3つの要素、特殊フォント、エディタ、解析ツール(テキストデータ、韻律解析、索引作成)が一通り揃ったことになる。

残りの解析ツールにおいてもWindows OSの独自の機能をを使用していないので、同じような手法により簡単に移植できると予測される。

現在、まだKylixのデレクトリパスの設定が不明瞭なため、実行ファイル単独での動作について確認していない。これは解析ツールのワープロ並の使用性を保証するものである。今後最優先的に吟味する予定である。

以上のようにKylixはその操作性が非常に優れていることが分かったが、残念ながら現時点では正常な動作はRed Hat Linux 7.2 迄しか保証されていない。それ以上のバージョンのRed Hat Linux に対して、正常に動作せず、開発元の今後の努力待ちの状態にある。

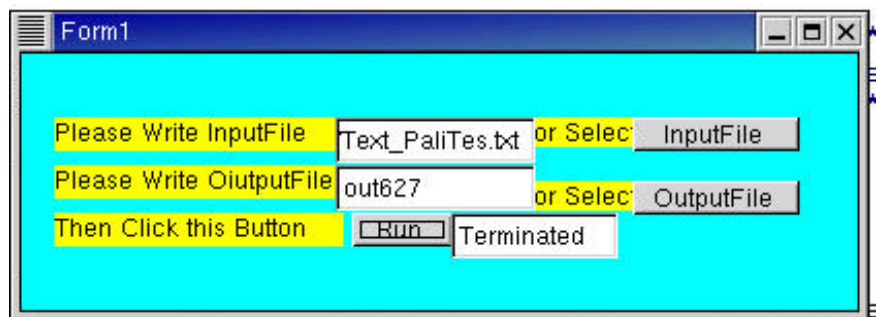


図4. 語彙索引作成ツールのダイアログウィンドウ

5. 討論

第3節で議論したように、OpenOffice.orgにおける特殊フォントのキーボード入力がまだ不自由な点がある。この点の対策として、(1) このキー操作のためのJavaアプレットTipsを開発し、OpenOffice.orgに組み込む、(2) またはEdisanをKylixに移植する、という2点について今後検討する。但し、Kylixではエディタ作成用のリッチエディットコンポーネントが無い。このコンポーネントの適用ファイルサイズは数GBと非常に大きいのに対し、代替りのメモコンポーネントは64KBに制限されているために、Edisanと完全に対応するエディタ作成はできない。この問題に対してどのように対処するかは今後の課題である。

またKylixに起因する問題点としては、このアプリケーションが多くのLinux Distributionに対応していないこと、又実行形式ファイルの動作については不安が残る。今後このアプリケーションのバージョンアップが期待される。

謝辞

この研究遂行に当たって適切な助言をしていただいた、与那嶺尚弘講師（仙台電波高専）、宮尾正大教授（室蘭工大）に深く感謝いたします。本研究は2003年度文部省・科学研究費補助金、基盤（B）(14310013)の助成の下に遂行したことを付記し、深謝致します。

参考文献

- [1] Y. Ousaka and M. Yamazaki, “Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer II”, *Literary and Linguistic Computing* (Oxford University Press 1996), Vol.11, No.1, pp. 9-17,
- [2] Y. Ousaka and M. Yamazaki, *Bhikṣuṇī-Vinaya: word index and reverse word index* (The Chuo Academic Research Institute, Tokyo 2002),

239pp., M. Yamazaki and Y. Ousaka, *A Word Index to Early Pāli Canonical Texts* (The Chuo Academic Research Institute, Tokyo 2000) 133pp.,

- [3] Y. Ousaka, M. Yamazaki and M. Miyao, *Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer with Object Files and Their Programs for Macintosh and Windows OS on CD-ROM* (The Chuo Academic Research Institute, Tokyo 2002), 86pp.,
- [4] <http://www.atmarkit.co.jp/flinux/rensal/linuxtips/321userttxfs.html>