

中期インド・アリアン聖典のデータベース
Database to the Canon in Middle Indo-Aryan

逢坂雄美*, 山崎守一*, 宮尾正大#

Yumi OUSAKA*, Moriichi YAMAZAKI* and Masahiro MIYAO#

* 仙台電波工業高等専門学校

〒 989-3124 宮城県仙台市青葉区上愛子

室蘭工業大学 工学部 電気電子工学科

〒 050 室蘭市水元町 27-1

* Sendai National College of Technology

Kamiyasi, Aoba-ku, Sendai 989-3124, Japan

Department of Electricity and Electronics, Muroran Institute of

Technology, Mizumotochou 27-1, Muroran 050, Japan

あらまし:

中期インド・アリアン諸語 (Middle Indo-Aryan) で書かれた古文献, 仏教 (パーリ語) とジャイナ教 (アルダ・マガダ語) 聖典の解析のための, 5つの計算機リソースを開発した. (1)アルダ・マガダ語又はパーリ語のローマ字化システム (Pali96 フォント) 構築. (2)Pali96 フォントによる電子化テキスト作成: ジャイナ教の重要な5つのテキスト (Uttarajjhāyā, Āyāraṅga, Sūyagaḍa, Dasaveyāliya, Isibhāsiyāim) 及び仏教の4聖典 (Dhammapada, Theragāthā, Therīgāthā, Suttanipāta) の収録. (3)聖典の韻律解析プログラム: 電子化テキストファイルが韻律解析の入力として使用される. (4)4つの索引 (語彙の正順索引・逆順索引とパーダの正順索引・逆順索引) の作成プログラム. (5)聖典の文法解析プログラム. 本論文では, フォント・電子化テキスト・解析ツール自体及び解析ツールを使用して得られた索引について議論する. 又, ウィンドウズ版特殊フォント用エディターについて述べる.

Summary :

We investigated automatic analysis of the canon in Middle Indo-Aryan by Personal Computer. Middle

Indo-Aryan is represented by literatures in Pāli and Prākṛit (called Ardha-Māgadhī), that is, the early Buddhist and Jainist canons. We constructed the following five computer tools: (1) making up the Prākṛit or Pāli fount set (Pali96 fount) in Roman; (2) preparation of the important Jainist texts and Pāli texts in Pali96 fount; (3) metre analysis program; (4) programs to compile four indexes, a word index, reverse-word index, pāda-index and reverse-pāda index; (5) grammatical analysis program. We discussed these computer tools, the calculated indexes obtained by using these tools, and editor tools for special fount Pali96 on Windows OS.

キーワード: 中期インドアリアン, アルダ・マガダ語, ジャイナ教, パーリ語, 初期仏教

Key words: Middle Indo-Aryan, Ardha-Māgadhī, Jaina, Pāli, early Buddhism

1. はじめに

中期インド・アリアン諸語 (Middle Indo-

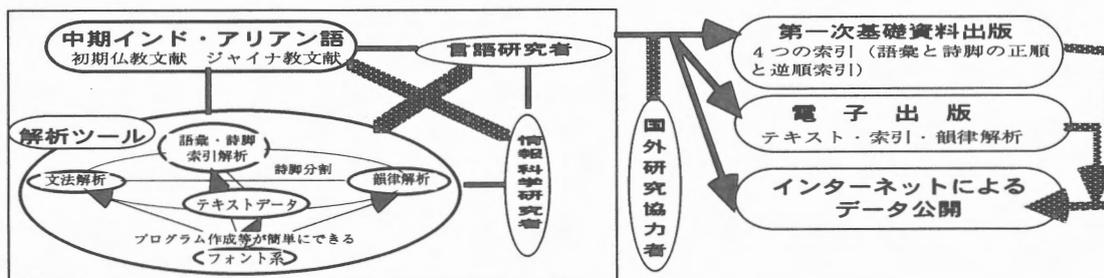


図1. 計算機リソースと研究組織

索引・逆順索引の作成プログラム。

V. 聖典の文法解析プログラム。

各ツールはそれ自体でも大変有益であるし、有機的にツールを使用することにより一層威力を発揮する。例えば、韻律解析は1つの半詩偈を2つのパーダへ分解することにも使うことができる。又、索引作成プログラムを使用して得られた語彙やパーダ索引により、テキストのチェックをできる。この結果得られた正確なテキストを他の解析に使用して、より正確な結果を得ることができる。このように、有機的で相互補完的な解析システムの構築は世界でも我々のものが唯一のものと思われる。

本論文では、これらの解析ツールを使用して開発された計算機リソース(フォントシステム・コンピュータブック・プレインテキスト・逆順索引)及び新たに開発した特殊フォント用エディターについて順に議論する(表1参照)。

表1. 我々のウェブサイトアップロードされたデータ。表中の記号○, ×は該当する計算機システムで使用可能なデータと使用不可のものを示している。

データ	マッキントッシュ	ウィンドウズ
特殊フォントPali96とキーボード入力法	○	○
ジャイナ教の5冊の聖典のコンピュータブック	○	×
ジャイナ教の5冊の聖典のプレインテキスト	○	○
初期仏教聖典の語彙の逆順索引	○	○
特殊フォントPali96用簡易エディターEDISANとその使用書	×	○

2. フォントシステム, Pali96

計算機解析に当たって、各刊行本をそのままコンピュータ処理することが出来れば、それが最良の方法である。しかし、デーヴァナーガリー文字(梵字、卒塔婆等に使われている文字)で書かれたテキストを、自動的に読み込むコンピュータツールが存在しないために、現時点ではこれではできない。デーヴァナーガリー文字テキストからローマ字化テキストへの変換には膨大なタイプ入力が必要され、多大な時間と経費が要求される。この作業中に、予期できないタイプミス等がローマ字化テキストに忍び込み易い。また、種々の索引を作成した場合、その構成には膨大な時間を要する。以上のことから、認識性の高い高品質の中期インド・アリアン諸語の為のローマ字化フォントを作成できるかどうか、研究を遂行する上で重要な役割を果たすことになる。さらに、適切に構築されたフォント体系はテキスト解析プログラムの作成を容易にする。これらの特性を持つローマ字化フォント(表2参照)を作成したが、全て1バイトコードに割り当てている。通常のローマ字フォントに無い特殊文字(例えば'm'等、表2参照)は、すべて拡張アスキーコード上に定義されている。

表2に、ウィンドウズPCにおけるフォントPali96のキー操作法と共に、アスキーコードへの割り当てを示している。このフォント系は、基本的にはマッキントッシュの場合と殆ど同一の入力法を採用している。つまり、マッキントッシュの Option key を、ウィンドウズPCでは Alt key に置き換えているだけであり、Shift key の使い方は全く同じである。また、アスキーコード割り当ては両方の計算機で同じにしている。このことにより、基本的には、マッキントッシュで作成したテキストは、ウィンドウズPCで

Aryan)で書かれた古文献は、仏教(パーリ語)とジャイナ教(アルダ・マガダ語)の聖典がその例に挙げられるように、世界の文化の貴重な宝庫であり、現代文明の重要な源流の一つである。初期仏教・ジャイナ教は厳格なカースト制度のパラモン教に対抗する形で同時期に発生したため、相互に関係を持ちながらその教理等を深化させてきた。それゆえ、今日では初期仏教を研究をする上でジャイナ教の研究は不可欠であり、その逆も真なりであると認識されている。しかしながら、これらの諸聖典は難解な言語で書かれていること、さらに質的・量的にも膨大であることもあって、今後、厳密な研究成果が待たれる分野である。

中期インド・アリアン諸語の文献の研究を飛躍的に進展させるためには、韻律解析・文法解析・語彙と構文論等の系統的な研究が不可欠である。テキストの韻律解析は批判的校訂本作成に不可欠であり、語彙の索引は翻訳に当たっての手助けとなる。また、語彙の逆順索引は文法構造解析に必要で、パーダ(詩脚)の正順または逆順索引はテキストの正しい読みの確立と並行詩脚を見いだすのに重要である。これらの解析に当っては大量のデータ、つまり多くのテキストを処理する必要があるが、幸いにもこの処理には現在のパーソナルコンピュータが適している。このような状況下で7~8年前から、ジャイナ教聖典の計算機解析についての研究を開始し、その後、仏教聖典についても議論を展開してきた。昨年はじめ、我々の研究概略と共に、開発した計算機ツールの概要・データベース等をインターネットのホームページに公開した(URL: <http://www.sendai-ct.ac.jp/~ousaka/>)。

近年、インド学・仏教学等において、インターネットによる情報発信・情報公開・情報交換が目覚ましい進展を遂げている。この分野での、著名なウェブサイトとして、Welcome Institute (Lodon) の Wujastyk (URL: <http://www.ucl.ac.uk/~ucgdkw/indology>) による "Indology - Internet Resources for Indological Scholarship" 等、多数存在する。これらのサイトには、パーリ聖典等の多数の電子化テキストやこれらのテキストを読むために必要なフォント等を含めて多種多様な情報が含まれており、当該分野の研究進展に寄与するものと期待され

る。

ところで、これらのウェブサイトを一見して気がつくことは、サンスクリット文献の研究に関するサイト(データベース [1])は非常に多く見られるが、それに比べてパーリ語関係のサイトはその数が少ないことである。さらに、アルダ・マガダ語文献に至っては、殆どほしいサイトがなく、我々のサイトの研究が際だっている。これは、各研究に対する興味のレベルを反映しているとも考えられるが、それ以上に各言語の研究の難しさの程度を正確に反映しているとも思われる。特に、日本では中期インド・アリアン諸語に関する専門家はほんのわずかしかないと言われている。その結果として、サンスクリット語の知識だけでは、当該言語に関する研究を正確に理解することができない、という認識さえなされていないのが実状である。

このような状況下にあつて、中期インド・アリアン諸語に関する正確な知識を有する国内外の研究者と情報科学の研究者との密接な共同研究により、我々はデータベースを作成することに成功した。これらの知見に基づき得られた最新のデータを元に構築された我々のデータベースは、それだけでも希少価値があると判断できよう。最近、海外の研究者から我々の開発したツールを使用しての共同研究の申し込みがあったが、これは開発したツールに対する評価が高まっている現れともいえよう。

これまで、我々はジャイナ教と仏教聖典の解析のための、下図のような5つの計算機リソースを開発した(図1参照)。

- I. アルダ・マガダ語又はパーリ語のローマ字化システム構築(これをPali96フォントと呼ぶ)。
- II. ローマ字化テキスト作成: ジャイナ教の重要な5つのテキスト(Uttarajjhāyā, Āyāraṅga, Sūyagada, Dasaveyāliya, Isibhāsiyāim)及び仏教の4聖典(Dhammapada, Theragāthā, Therīgāthā, Suttanipāta)をPali96フォントを使ってマッキントッシュに収録した。
- III. 聖典の韻律解析プログラム: 電子化テキストファイルが韻律解析の入力として使用される。
- IV. 語彙の正順索引・逆順索引とパーダの正順

も同一コードで表現できることになる。実際、ウインドウズPC用のジャイナ教プレインテキストは、マッキントッシュ用テキストをウインドウズPC用に変換して、作成している。

表2. ウインドウズPC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。記号<A>,<S>はそれぞれAlt key, Shift keyを示している。正規の文字は通常通りにタイプできる。特殊文字はAlt key, Shift keyと通常のキーを組み合わせることにより入力できる。第3列と6列の数字はアスキーコードを表している。

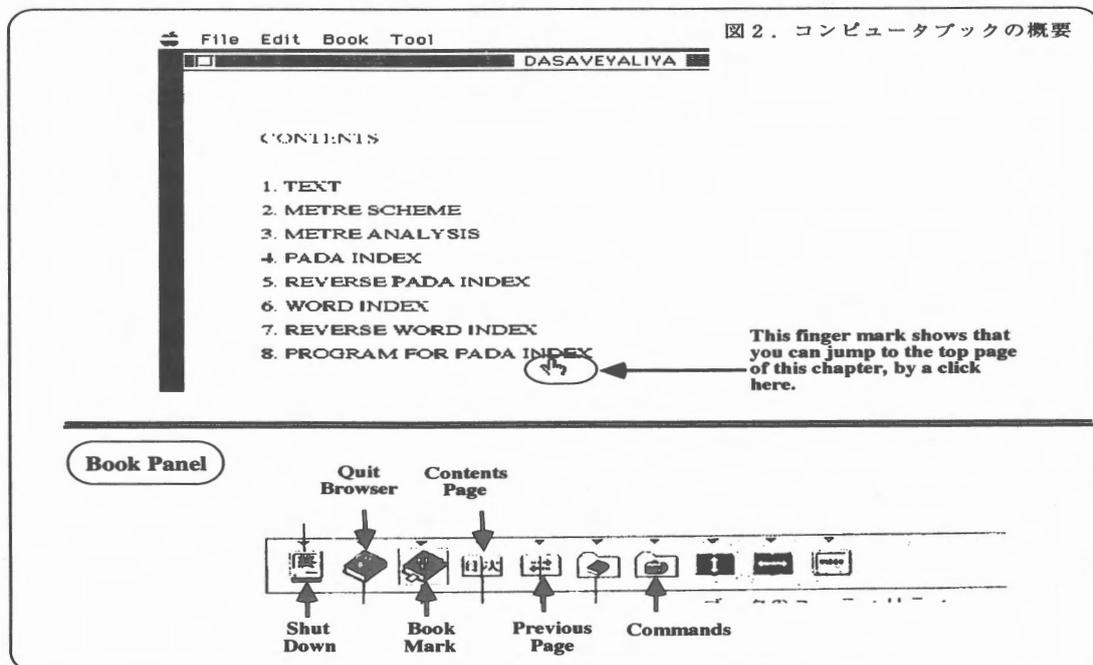
ā	<A>+n	136	t	<A>+t	160
ā	<A>+a	140	th	<S>+<A>+y	231
ī	<A>+<S>+'	147	d	<A>+d	182
i	<A>+i, i	148	dh	<A>+i, U	243
ū	<A>+e, u	156	ṇ	<A>+n, n	150
ū	<A>+u, u	159	th	<A>+y	180
m	<A>+m	181	dh	<A>+z	189
kh	<A>+k	251	ph	<A>+p	185
gh	<A>+g	187	bh	<A>+b	186
ṅ	<A>+g	169	l	<A>+l	194
ch	<A>+c	141	!	<A>+<S>+!	241
jh	<A>+j	198	ś	<A>+s	167
ñ	<A>+<S>+i	246	ṣ	<A>+x	197

マッキントッシュとウインドウズPC用のフォントは共にTrue Typeフォントの形式で作成され、我々のウェブサイトから各自の計算機にダウンロードすると、自動的に解凍される。その後、各自の計算機に応じてインストールすればよい。

3. コンピュータブックとプレインテキスト

ジャイナ教の重要な5つのテキスト(Uttarajjhāyā, Āyāraṅga, Sūyagaḍa, Dasaveyāliya, Isibhāsiyāim) 及び仏教の4聖典(Dhammapada, Theragāthā, Therīgāthā, Suttanipāta)をPali96フォントを使ってマッキントッシュに収録した。仏教の4聖典の電子化テキストはパーリ文献協会の前会長ノーマン氏より提供されたものであり、ノーマンフォントで作成されている。このテキストを我々のフォントで書き換え、索引作成に活用できるようにした。著作権に抵触するため、電子化テキストは残念ながら公表することができない。

一方、ジャイナ教の聖典には著作権に関する制限がないので、電子化テキストのみならず、図2に示すように、韻律解析結果・詩脚の正順索引と逆順索引・語彙の正順索引と逆順索引・詩脚の索引作成プログラムを含めて、各聖典ごとにコンピュータブックを作成した。言語研究上必要な基礎資料がテキスト資料と共に、1冊のコンピュータブックにまとめられているので、非常に有効な電子化ツールである。図2の下半分に、コンピュータブックの操作パネルを示している。このコンピュータブックもフォントと同様、我々のウェブサイトから自由にダ



ウンロードして使用できる。以下に各項目の特性について簡単にふれる。

(1) 電子化テキスト作成とブレインテキスト提供：

ジャイナ教聖典に関しては次のようにテキストを作成した。最初、原典テキストからパソコン入力をし、入力ミス等のチェックをした。その後、韻律解析プログラムを使用して詩を分析し、その結果に基づき詩脚に分割した。そして、各テキストごとに詩脚索引を作成し、得られた結果を言語学的に厳重にチェックした。その索引を出版して、国内はもとより国外の当該研究者に配布した。その後、今度は詩の部分だけではなく散文の部分を含めて語彙の索引を作成し、同様に配布した。この段階でも残存するテキストミスを修正している。最終的に、5冊の聖典群とこれまでの批判的校訂本を含めての詩脚の総索引を作成し、出版公表した。この段階でも残っていた若干のミスを修正した。このように、種々の言語学的知識を総合的に活用して、極力ミスのない電子化テキストを作成している。幸い、我々の作成した電子化テキストについてのエラーの指摘が今のところ一件もない。なお、初期仏教聖典のテキストも同じような手順を踏んで作成している。結果として、正確な電子化テキストを種々の解析に利用できるようになった。

このように作成された電子化テキストを、コンピュータブック特有の書式に基づき収録していると同時に、論文作成等の便宜のためにブレインな形でのテキストの提供も行っている。

(2) 韻律解析：

中期インド・アリアン諸語の韻律解析のために、我々が採用している韻律規則・手法およびそのプログラミング法を、専門誌 *Literary and Linguistic Computing* (LLC, オックスフォード大学出版) に投稿した二つの論文で詳しく議論している(参考文献15と16参照)。最初の論文は、中期インドアリアン諸語の文献研究の専門家ではなく、サンスクリット語の研究者が審査に当たったらしく、論文で議論している韻律解析は何ら新しいものがなく、既にサンスクリット語

に関連して開発されているので、掲載の価値はないとの講評を受けた。その後、中期インド・アリアン諸語の知識を有する審査員からLLCに掲載するだけの価値があるとの論評があり、掲載が許可されることになった[2]。第二の論文では第一の論文でまだ議論されていなかった韻律の残りの部分を議論し、それらを使って得られる解析結果を援用して、言語学的にどのような新しい知見が得られるかについて、詳しい議論を展開している。

韻律の研究には、まず最初に半詩偈(下図参照)ごとの韻律図式(音節パターン)を計算しなければならない。

parisahānam pavibhatti kāsavenam paveiyā /	← 第1半詩偈
taṃ bhe udāharissāmi ānupuvīm suṇeha me //	← 第2半詩偈
1 digimchāparigae dehe tavassi bhikkhu thāmavam /	
na chinde na chindāvac na pae na payāvac //	

つまり長母音(‘-’)と短母音(‘◌’)配列を求めなければならない。これは次の韻律規則に基づきなされる。(a) 全長音は、いつも長母音として読まれる。(b) 短母音の後に2つの子音を伴う場合は長母音として読まれる。(c) 特殊文字‘ṃ’を後に伴う短母音はいつも長母音として読まれる。(d) 上の(b)と(c)の条件以外の短母音は常に短母音として読まれる。なお、上記の(b)と(c)の計算において、語彙間の空白を無視することを注意しておく。次に、韻律表示に必要な音節数とモーラ数について説明しておく。音節数は長母音と短母音の総数を表わし、モーラ数とは短母音を1とし長母音を2と数えた時の総数を示す。

表3に示すような中期インド・アリアン諸語の韻律(3大韻律, Akṣaracchandā, Mātrāchandā, Gaṇacchandā)に基づき解析を行った。表3aに示されたAkṣaracchandāは、音節数によって分類される。Ślokaは半詩偈に16音節を含み、TriṣṭubhとJagatīはパーダ(詩脚=1/2半詩偈)にそれぞれ11, 12音節を含む韻律である。表3bの上部に示されたMātrāchandāは、パーダに含まれるモーラ数によって規定される。半詩偈のモーラ数に換算して、28~30のものをVaitāliya, 30~34のものをAupacchandāsakaと呼ぶ。表3bの下部に示され

たGaṇacchandāsは、半詩偈中のモーラ数によって規定され、その典型的な韻律、Āryāのモーラ数は、26, 27, 29, 30である。これらの韻律はすべて、図に示されるような特有な韻律図式(スキーム)を持っている。

我々は、韻律解析プログラムを使って各詩偈の韻律パターンを求め、表3のスキームと比較し各詩偈の韻律種類を決定した。韻律割り当てが、1つだけでなく複数個が可能な場合には、それらをすべて表示し、正しい韻律選択は言語学の専門家が決定するようにしている(この表示法をAlternative Representationと呼ぶ)。標準的な韻律スキームで分類できない詩偈には、このAlternative Representationによる表示が現れることになる。この表示が現れる箇所が問

表3a

韻律名	韻律図式	音節数又はモーラ数
Śloka	uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu uuuu	半詩偈の音節数 = 16
Modified Śloka	接頭辞を考慮に入れて、17音節を16音節に変換(韻律図式はシュローカと同じ) 連続する2短音を1長音に変換(韻律図式はシュローカと同じ)	半詩偈の音節数 = 17
Triṣṭubh	下記の3グループの総ての組み合わせ Opening (uu-; -uu; -uu-; -uu-), Break (-uu; -uu; uu; uu; uu-; -uu) Cadence (-uu; -uu-)	バーダの音節数 = 11
Jagati	下記の3グループの総ての組み合わせ Opening (uu-; -uu; uu-; uu-; uu-; uu-), Break (-uu; -uu; uu; uu) Cadence (-uu; uu-; uu-; uu-; uu-; uu-)	バーダの音節数 = 12
	半詩偈で両韻律の4個のバーダ組み合わせが可能 (Modified Ślokaと同じ修正を受けたものが上記半詩偈の韻律に含まれている。)	半詩偈の音節数 = 22~25

表3b

韻律名	韻律図式	音節数又はモーラ数
Vaitāliya	(Odd pāda) (Even pāda) u -uu- u-uu u- -uu- u-uu	バーダのモーラ数 = 13~16
Aupacchandasaka	u -uu- u-uu u- -uu- u-uu	= 14~18
	半詩偈で両韻律の4個のバーダ組み合わせが可能	半詩偈のモーラ数 = 26~34
韻律名	韻律図式	音節数又はモーラ数
Āryā	uuu uuu uuu uuu u uuu u uuu uuu uuu uuu u uuu uuu uuu uuu uuu u uuu uuu uuu uuu uuu u uuu	半詩偈のモーラ数 = 26, 27, 29, 30
Old Āryā	uuu uuu uuu u - uuu uuu uuu uuu uuu uuu u- uuu uuu uuu uuu uuu uuu u- uuu uuu uuu	半詩偈のモーラ数 = 28~31

題となる詩偈であるので、この韻律結果を活用して言語学者は研究をより効率的に推進できるであろう。

このように開発されたツールを使って得られた韻律解析結果をコンピュータブックに収録し、韻律関係に深い知識を有するヨーロッパの研究者及び現存するジャイナ教徒のアルダ・マガダ語の専門家等に配布したが、好評である。

(3)詩脚と語彙の索引：

我々は、詩脚の正順索引・逆順索引と語彙の正順索引・逆順索引の作成プログラムを自作し、このプログラムと作成した電子化テキストを使用して索引を求めた。この4~5年の間に十数冊の本を出版している(詩脚の逆順索引については図3参照。参考文献参照)。ジャイナ教聖典については、そのデータをコンピュータブックに収録している。前に述べたように、これらの基礎データは当該研究にとって重要な役割を果たす。つまり、語彙の正順索引は翻訳に当たっての手助けとなり、語彙の逆順索引は文法構造解析に必要で、パーダ(詩脚)の正順または逆順索引はテキストの正しい読みの確立と並行詩脚を見いだすのに重要である。日本からこのような基礎的で重要な研究資料が相当量公開されることに、国内はもちろん、海外の研究者からは驚きの目で見られるとも非常に好評である。

atha maṃ bhaṇati tāto idh	430a
pabbajjā dukkarā puttaka	463d
nihato tvam asi antaka	59d 62d 142d 188d
patinā khujjakena ca	195d 203d 235d
annena ca pānena ca	11d
annena pānena ca	429a 430c T3-409a
sobhaggena yasena ca	409a T4-429a T4-430c
kāyena ca vācāya ca	409a
kāyena vācāya ca	72b
kāṇāya ca khañjāya ca	T3-452c T4-452c
mā evaṃ putta avaca	452c
ehi bhadde ti avaca	438c T4-438c
jātiyā maraṇassa ca	415a
sassūyā sasurassa ca	109c
sassuyā sassurassa ca	T3-109c
kisā paṇḍu vivanna ca	165b

図3. 詩脚の逆順索引。右側の数字等は詩脚番号等を示している。

4. 初期仏教聖典の索引

これまで『ダンマバダ』、『ヴィナヤ』、『ディー

ガ・ニカーヤ」等を含め、初期仏教文献の重要な原典テキストはパーリ文献協会(PTS, オックスフォード)より計画的に発行され、研究者に多大の便宜を供してきた。しかしながら、これらの原典テキストには、語彙索引が殆ど付されていないのが現実であった。この作成に当たっては、パーリ語に関する深い知識とそれに基づく処理プログラム作成能力の両方が必要であったが、これまでこの2つを同時に満たす研究組織がなかったため、これまで索引が作成されなかったのが実状であった。我々は、言語学の研究者(ケンブリッジ大学のノーマン教授とコーン博士及び山崎)と情報科学の専門家(逢坂・宮尾)との密接な共同研究(図1参照)により短期間で膨大な索引を完成させている。

これら索引作成に当たって、『ダンマパダ』はノーマン教授の作成による電子化テキストを使用している。『ヴィナヤ』と『ディーガ・ニカーヤ』については、タイのダンマカヤファウンデーションで作成したものを、PTS(ノーマン教授)経由で提供され、使用している。後者のテキストは、原典となったPTS版テキストの誤植・エラー・多数のタイプミス等を含んでいるので、可能な限りそれらすべてを訂正している。計算機による索引作成をし、それに基づき言語学者がチェックする、という操作を数回繰り返して、正確な索引を完成している(参考文献参照)。この結果として、正確な電子化テキストを作成できた。このように、現代の最高水準の言語学研究に基づき、テキストチェックを実行している。

我々は上記の手順で、『ダンマパダ』の4つの索引を求め、1995年に出版した。この索引は好評で、地味なデータブックにもかかわらず、既に増刷されている。『ヴィナヤ』、『ディーガ・ニカーヤ』についてもすでに語彙の正順索引をPTSより出版している。『ジャータカ』については、その索引を作成中である。これらの研究の成果の1つとして、山崎は”PTS版テキストの限界—『ディーガ・ニカーヤ』を中心に”[3]の論文でPTS版の欠陥・種々の問題等について議論している。これは、PTSで出版しているテキストには何らエラーがなく信頼できるという、誤った認識を打破する貴重な論文である。この論文の概要を、我々のウェブサイトに掲載している。パーリ語文献に比べて、より難解なアルダ・マガダ語文献には、より一層精密で慎重な研究が要求される。

我々は、上記の手順を踏んで極めて正確に作成

された電子化テキストを基に、『ヴィナヤ』と『ディーガ・ニカーヤ』の語彙の逆順索引を作成し、ウェブサイトに掲載した。当初、マッキントッシュ用の索引だけを掲載していたが、フランスの研究者からの要望により、この索引をウィンドウズPCでも使用できるように拡張した。この過程で、ウィンドウズPC用のフォントを作成した。この索引のダウンロード状況から判断して、国内の研究者もよく使用しているようである。又、イギリスのL. S. Cousins教授から、すでに作成済みの『Mahaaniddesa』の電子化テキストを基に、我々のツールを使用して、語彙の逆順索引を作成してはどうだろうか、との申し込みがあった。今後、この提案に基づき、当該テキストの索引を作成し、ウェブサイトに掲載する予定である。なお、『ヴィナヤ』等の語彙の正順索引は、著作権の問題があるのでウェブサイトへの掲載を見合わせている。

5. ウィンドウズ版特殊フォント Pali96用簡易エディター

第1節で述べたように、近年、インド学・仏教学等においてもインターネットによる情報発信・情報公開・情報交換が目覚ましい進展を遂げている。多数のウェブサイトには、パーリ聖典等の多数の電子化テキストや、これらのテキストを読むために必要なフォント等を含めて、多種多様な情報が含まれている。

ところで、これらの情報を活用しようとする問題点に遭遇する。例えば、特殊フォント(表2参照)で書かれたテキストをダウンロードし、ウィンドウズPC上で、MSワード等を使って開くと、いわゆる文字化けを起し、正確に読むことができない(図

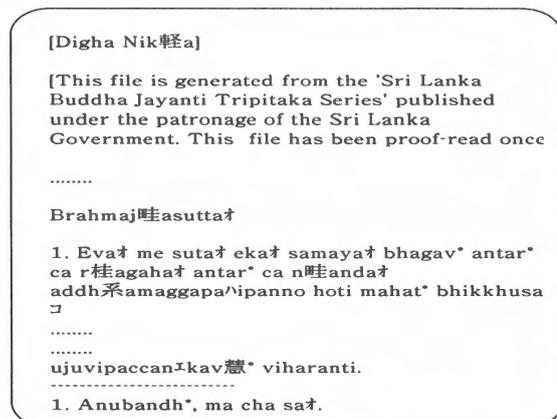


図4. 文字化けしたテキスト

4参照)。これは、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)上で定義された特殊なフォントを使用する場合に、現われる特有な現象である。この現象は、拡張アスキーコード上で定義された特殊フォントを、2バイトコードのフォントの取扱を標準にする日本語ワードプロセッサで表示する場合、その機能にバグがあることに起因する、と考えられる。又、現在のところ、簡便なキーボード入力法もない。

一方、KH一方式は127番目までの通常のアスキーコード上で展開された英字フォントの特殊な組み合わせで「ā」等の特殊フォントを表現している。通常のアスキーコードのみを使ってテキストを表わしているような場合、文字化け問題は出現しない。然し、KH一方式よりは、印刷されたものと画面上の表示が一致する、所謂 WYSIWYG (=What you see is what you get.) 形式で表現されたテキストは分かりやすいので、言語学的な研究の能率も向上し、又、語彙の索引作成等にも適している、という多くの利点がある。さらに、国内外で1バイトの特殊フォントを基に作成された貴重な基礎資料を有効に活用するという観点から判断して、文字化けすることなくテキストの入出力を正確に実行できるエディターを開発することは、当該研究の進展に多大な寄与をすることになる。

よく使われる1バイトの特殊フォントとしてはノーマンフォント及び我々のPali96フォント等がある。これらのフォントには、ウインドウズPC用のTrue Typeフォントが共に用意されている。

文字化け現象は以下のようにして解消した。

フォント自体が1バイトコードで作成されているので、半角英数にモードを切り替え、その後フォントをノーマン等にして、ファイルの読み込みを実行した。しかしながら、文字化け現象をどうしても解消できなかった。一方、電子化テキストを入力テキストとして、我々の作成した語彙索引プログラムを実行し、出力結果の語彙の表示を見ると、「ā」等の特殊文字群を正確に表示することができず、フォント割り当てができないことを示す「?」の表示がなされていた。

以上のことから、ウインドウズPCでは、拡張アスキーコード領域内で定義された特殊フォントが、正常にインストールされていない可能性がある、と考えられる。しかし、ウインドウズPCのプログラム項目のAccessaryにあるキーコード表をクリックしたところ、フォントが正しくインストールされていることが分かった。つまり、フォント自体には何らの異常もなかった。

次に、プログラムソフトDelphiに標準装備してあるテキストエディタで作成したテキストファイルを使用して、語彙索引プログラムを実行させたところ、「ā」等の特殊文字を含む単語を正確に出力できた。プログラムソフトDelphiはテキスト処理をするときには、当然のことながら拡張アスキーコードを扱うことになるので、それらのフォント管理情報を正確に処理しており、正常な表示が得られたと考えられる。

このことから判断して、MS Word、Word Pad等のワープロソフトには、日本語における半角カナが割り当てられている拡張アスキーコード領域で定義

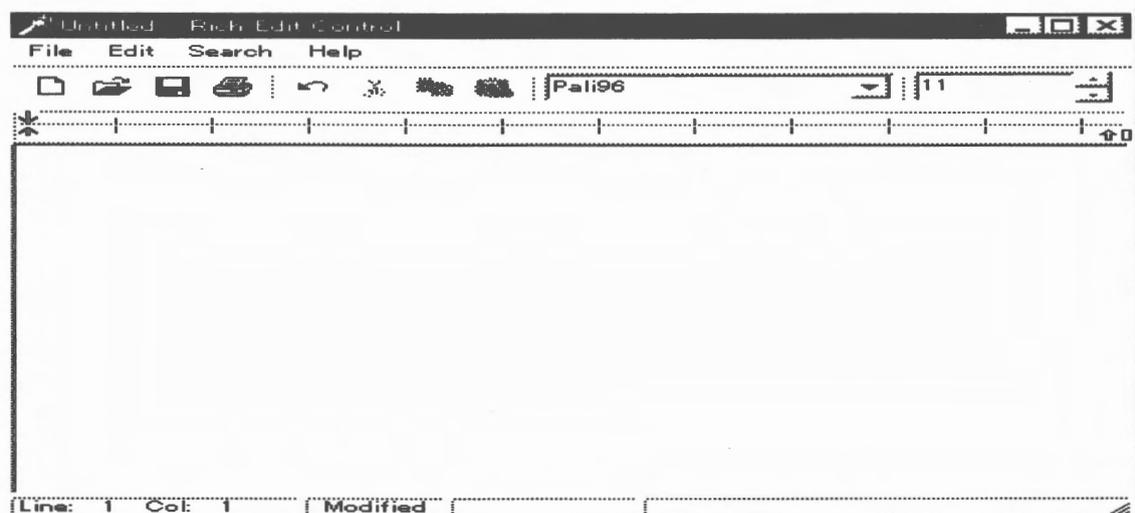


図5. 簡易エディターEDISAN.

されているフォント表示の管理に、バグがあると結論づけられる。

それ故、プログラムソフトDelphiを使って電子化テキストを処理することにより、拡張アスキーコード上で定義されたフォントを、正確に表示できる可能性があると考えられる。拡張アスキーコード上で定義されたフォントを使って作成されたテキストデータを、クリップボード経由でDelphiのテキストエディタにコピーすると、正しい拡張アスキーコードの管理情報が保持され、正確な表示ができることが分かった。これのこをベースにして、特殊フォントを使った場合でも正確なテキスト表示の処理をするエディター EDISAN を Delphi で作成した。ウインドーズPCの場合、マッキントッシュと違って、これらのフォントの簡便なキーボード入力法がなかったので、その入力法をも開発し、エディターに付加してある(表2参照)。

図5にエディターの構成画面を示している。通常のワープロソフトの画面とほぼおなじであり、使用方法も殆ど同じである。このエディターも我々のウェブサイトに掲載してある。

6. 討論

解析ツールのうち韻律解析についてはまだ不十分な点があり、改良の余地がある。また、仏教文献の Suttanipāta とジャイナ教文献の Āyāraṅga については、計算機解析の点からみて、議論すべき点はまだ残されている。ここでは以上のことについて順に議論する。

韻律解析について。

(1)詩偈のパッチワーク：

我々の解析ツールでは、音節数が19～21の半詩偈の場合そのモーラ数が、半詩偈のモーラ数で規定される韻律 Āryā 等に合致しない場合には、表3のどの韻律にも分類できずに、Examine という表示が出力される。しかし、これらの音節数の場合は、パーダの音節数が8の Śloka, 11の Triṣṭubh, 12の Jagatī の適当な組み合わせと、Modified Śloka 等で考慮した2短音を1長音に変換する処理を併せて考慮することにより、解析できる可能性がある。中期インド・アリアン聖典の韻律では、いわゆるパッチワーク的な組み合わせが現れることから判断して、上記の組み合わせを検証する価値がある。これまでのプログラ

ム処理では半詩偈の処理を基本にしていたが、これらの音節数の詩偈に対してはパーダを基本構造にすることにより、解析できると予想している。

又、モーラ数(及び韻律スキーム)によりパーダが規定される Vaitāliya と Aupacchandāsaka についても、同様にパッチワーク的韻律解析処理を考慮する必要がある。

(2)不完全な詩の解析：

仏教文献の Suttanipāta とジャイナ教文献の Āyāraṅga には半詩偈構成が不完全なもの、パーダが欠落したもの、等がみられる。これらの不完全な詩偈の処理は現在のプログラムでは不可能であるが、(1)の詩偈のパッチワーク処理を考慮することにより、解析できる部分があるのではと予想している。さらに、これら2冊の文献には定型的な韻律で解析できないものが多いが、これらもパッチワーク処理を考慮したプログラムにより相当数解析できる可能性がある。

(3)従来の解析プログラムとの比較：

他の韻律解析プログラムは、殆どサンسكريット語をベースにしている。今まだ述べてきたように、中期インド・アリアン聖典の韻律を理解するのはサンسكريット語の知見だけでは不十分である。つまり、Modified Śloka 等で考慮した2短音を1長音に変換する処理とか、(1)で述べた詩偈のパッチワーク的処理等を考慮する必要がある。より正確に言うと、(1)の詩偈のパッチワーク的処理は古典インド叙事詩の解析でも考慮する必要がある。中谷氏は、マハバータ[4]の韻律解析を実行し、Mātrāchandas をピックアップしている。しかしながら、パッチワーク処理を考慮していないようで、簡単なチェックによると、例えば、詩節番号12.187.59, 60の詩偈をリストアップしていないことが分かる[5]。この詩偈は、パッチワークで構成された詩偈で、Vaitāliya の偶数パーダの韻律スキームが奇数パーダに現れてきている。我々のプログラムでは、まだこの処理を取り込んでいないが、Alternative Representation の表示をするので、要注意の韻律であることを容易に判断できる。なお定型的な Vaitāliya の韻律は、我々のプログラムでも当然判定できる。このほか、我々の処理では2短音を1長音に変換する処理を考慮していることから、より適用範囲の広い解析プログラムといえよう。

次に、仏教文献の *Suttanipāṭa* のテキストで、上記の(2)で指摘したこと以外の問題点を指摘しておく。ドイツの著名な専門家ボレーがこのテキストの詩脚の順列索引を出版している (*The pādas of the Suttanipāṭa, Studien zur Indologie und Iranistik, Monographie 7 (Reinbek 1980)*)。ボレーは通常、散文と考えられている箇所を詩脚索引にリストアップしている (例えば p. 46 *tassa vagassa uddānam* p. 71, 1.2, 等を参照)。又、詩脚の欠けているもの、2つまたは3つの詩脚がまとまってしまう分離できない部分もある。これまで既にこのテキストの全詩形が確定している、と思われている向きもあるが、これまで議論してきたことから判断して、明らかにこのテキストの韻文のすべてが特定できているとはえないであろう。詩脚の順列索引のみが出版されていること、一部の詩脚が索引から落ちていること等から、数冊の批判的校訂本からの詩脚のデータを含めて、*Suttanipāṭa* の正順索引と逆順索引を我々の開発したツールを使って求め、出版する準備をしている。

謝辞

当該研究遂行に当たり、プログラム作成等で協力していただいた電波高専卒業生 (林礼子、門間薫、善生良二、瀬戸紀子、宮澤綾子、小野寺里美、平野健嗣の諸君) に感謝いたします。この研究は文部省・科学研究費補助金 (特定領域研究「人文科学とコンピュータ」) の助成の下に遂行したことを付記し、深謝致します。

注

- [1] John Smith's Cambridge archive (Wujastyk のウェブサイト (<http://www.ucl.ac.uk/~ucgadkw/indology>) にリンクされたサイト) 等を参照。
- [2] この事例は、中期インド・アリアン諸語の聖典に関する研究論文の審査には、サンスクリット語だけでなく当該言語に関する正確な知識を有する研究者による慎重な評価が要求される、ことを端的に示している。
- [3] 山崎、PTS版テキストの限界—「ディーガ・ニカーヤ」を中心に—、「仏教研究」第27号 (平成10年3月) p. 137。
- [4] [1] のウェブサイトからもデータをダウンロードできる。John Smith's Cambridge archive には Kyoto file の edited version が掲載されている。言語の計算機解析には、極力正確な電子化テキストを使用することが重要であるので、このことを付記しておく。
- [5] 中谷英明、インド古典二大叙事詩の韻律、文部省科学

研究費補助金・重点領域研究「人文科学とコンピュータ」1995年度研究成果報告書、p. 213。本報告所に記載の徳永氏の論文を参照のこと。

参考文献

1. M. Yamazaki, Y. Ousaka and M. Miyao, *Dasaveyāliya: Pāda Index and Reverse Pāda Index*, Monograph Series 1 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1994.
2. M. Yamazaki and Y. Ousaka, *Isibhāsiyāim: Pāda Index and Reverse Pāda Index*, Monograph Series 2 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1994.
3. M. Yamazaki and Y. Ousaka, *Āyāraṅga: Pāda Index and Reverse Pāda Index*, Monograph Series 3 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1994.
4. M. Yamazaki and Y. Ousaka, *Sūyagaḍa: Pāda Index and Reverse Pāda Index*, Monograph Series 4 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1995.
5. M. Yamazaki and Y. Ousaka, *Uttarajjhāyā: Pāda Index and Reverse Pāda Index*, Monograph Series 5 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1995.
6. M. Yamazaki and Y. Ousaka, *A Pāda Index and Reverse Pāda Index to Early Jain Canons*, Kosei Publishing Co., Tokyo 1995.
7. M. Yamazaki and Y. Ousaka, *Dasaveyāliya: Word Index and Reverse Word Index*, Monograph Series 6 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1996.
8. M. Yamazaki and Y. Ousaka, *Isibhāsiyāim: Word Index and Reverse Word Index*, Monograph Series 7 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1996.
9. M. Yamazaki and Y. Ousaka, *Āyāraṅga: Word Index and Reverse Word Index*, Monograph Series 8 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1996.
10. M. Yamazaki and Y. Ousaka, *Sūyagaḍa: Word Index and Reverse Word Index*, Monograph Series 9 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1996.
11. M. Yamazaki and Y. Ousaka, *Uttarajjhāyā: Word Index and Reverse Word Index*, Monograph Series 11 of *Philologica Asiatica*, The Chūō Academic Research Institute, Tokyo 1997.
12. M. Yamazaki, Y. Ousaka and M. Miyao, *Indexes to the Dhammapada*, The Pali Text Society, Oxford 1995.
13. Y. Ousaka, M. Yamazaki and K.R. Norman, *Index to the Vinaya-piṭaka*, The Pali Text Society, Oxford 1996.
14. M. Yamazaki, Y. Ousaka, K. R. Norman and M. Cone, *Index to the Dīgha-nikāya*, The Pali Text Society, Oxford 1997.
15. Y. Ousaka, M. Yamazaki and M. Miyao, "Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer," *Literary and Linguistic Computing*, Oxford, vol. 9, no. 2 (1994) pp. 125-36.
16. Y. Ousaka and M. Yamazaki, "Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer II," *Literary and Linguistic Computing*, Oxford, vol. 11, no. 1 (1996) pp. 9-17.