

インド・アーリヤ語古文献データベース IAMDB の評価

逢坂雄美

仙台電波工業高等専門学校 総合科学科

我々は言語学研究者と情報科学研究者による共同研究により、実用的な中期インド・アリアン文献データベース (IAMDB) を構築してきた。その中核をなす文献解析システムは、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。文献解析システムを活用して作成した多数の重要な文献の語彙・詩脚索引群を、提携している国内外の研究機関を通して公表・出版している。本論文では、電子化テキストと使用解説付き解析ツールに焦点を絞って議論する。

Evaluation of IAMDB Database for study of Canons in Middle Indo-Aryan

Yumi Ousaka

Faculty of General Sciences

Sendai National College of Technology

We have developed a practical database for study of the canons in middle Indo-Aryan (MIA), in collaboration with linguists and computer scientists. The analysis system for canons plays a dominant role in this database, which comprises three ingredients such as the special font in MIA, editor, and analysis tools (text data, metre analysis programs and index production programs). So far we have published about 30 indexes to the important manuscripts from two institutes in Japan and in England. We concentrated on the discussion on the preparation of the electronic texts and the analysis tools.

まえがき

中期インド・アリアン (B.C. 5-A.D. 10) の古文献は、仏教（パーリ語）・ジャイナ教（アルダ・マガダ語）・仏教混淆梵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。これまで、これらの古文献は難解な言語で書かれており、異なる歴史的背景と異なる地域性を持つ文化圏で長期間に亘って作成されたこと、さらには質的・量的にも膨大であり多様な文献群が存在することにより、十分な研究がなされていない。ジャイナ教文献に批判的校訂本が少ないのにに対し、『ヴィナヤ・ビタカ』等を初め、初期仏教文献の重要な原典テキストは組織的に研究され、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。これらのテキストは校訂上の方針の誤りが指摘されており、近年再校訂の必要性が議論されるようになった。又これらの原典テキストには基礎資料となる語彙索引が殆ど付されていないのが現実であり、付されている場合でも不完全で誤りが多く、完全な索引が研究者に切望されている。この目的達成に向けて、90年代当初から我々は言語学研究者と情報科学研究者との共同研究の下に、パーソナルコンピュータを使用して、古文献を系統的に解析できるツールを開発してきた。

この解析ツールを、始めにマッキントッシュ上で、次いでウインドウズ、及びLinux上で開発してきた。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 [1]。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、PTSの共同研究者と協力して、この数十年間待望されていた『ダンマパダ』の語彙の正順・逆順索引と詩脚の正順・逆順索引、『ヴィナヤ・ビタカ』等の語彙索引を、短期間で作成・出版している。これまで、30冊ちかくの基礎資料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]。Web上のデータ及び公刊した書籍群 [2] は、広い意味でインド・アーリヤ語古文献のデータベース (IAMDBと呼ぶ) を構成していると言えよう。IAMDBは、パーリ語辞書編集の研究・学術的基礎研究等に対して多大な寄与をすることになり、印度学・仏教学研究を格段に発展させることができると期待される。

我々の作成したPali96フォント（表1参照）がデータベース (IAMDB) の基本要素であり、中期インド・アリアンに属する主要言語で共通に使用できる。このフォントはTrue Type フォントであり、マッキントッシュ (OS9, OSX), ウインドウズ (OS 98, OS XP), リナックスの異なるプラットフォーム上で共通のコードの元に使用できる。但し、リナックスでは各Distributionにより若干の使用上の難易度があるが、その環境整備が強力に進められている現状を考慮すると、ごく近い将来解消されると期待される。

国内外において、電子化テキストは主にNormy

フォント（元PTS会長ノーマン先生作成）に基づき作成されている。Pali96フォントはNormynフォントと殆ど等価であり、計算機解析により適している。Normynフォントで書かれたテキストをPali96フォントによって書き換えることは簡単にできる。ノーマン先生から提供されたNormynフォントで作成されたパーリ語文献の電子化テキストをこのように処理し、計算機解析に使用している。

次節の電子化テキストの項目で述べているように、当該分野では多くの古文献の電子化テキストが、Normynフォントで作成され、研究進展に寄与している。又、日本国内における当該分野の諸学会誌では、NormynフォントまたはPali96フォントが英字フォントの基準として採用されている。このように、現状ではユニコードベースのフォントでなくTrue Type フォントが主流となっている。現時点では、ユニコードベースのフォントを主体とするワープロが、当該分野の言語学者にはまだ普及していない。正確に言えば、これらのフォントも又使える環境を備えたワープロが提供されているが、今後徐々に普及するのであろうが、今ではまだ使用の主流ではない。このように、現状で

表1. パーリ語ローマ字とPali96フォントのウインドウにおけるキー操作法・アスキーコード割り当て。
a, i, u, e, o は長母音であり、a, i, u は短母音である。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 “ā” 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。同様にして他の特殊文字 “m”, “kh” 等を打つことができる。特殊文字 “ī” は Alt key と 正規の文字 “i” を同時に押し、その後で引続き文字 “ī” を打つことによりタイプできる。同様にして “ū” を打つことができる。最後に、特殊文字 “īb” 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。

順序	文字	入力法	コード	順序	文字	入力法	コード
1	a	a	97	29	j	j	106
2	ā	<O>+a	140	30	jh	<O>+j	198
3	i	i	105	31	ñ	<O>+<S>+i	246
4	ī	<O>+i, i	148	32	!	<O>+t	160
5	u	u	117	33	þ	<O>+<S>+y	231
6	ū	<O>+u, u	159	34	đ	<O>+d	182
7	r	<O>+r	168	35	đh	<O>+i, U	243
8	ř	<S>+<O>+m	229	36	ñ	<O>+n, n	150
9	e	e	181	37	t	t	116
10	ai	<S>+<O>+g	101	38	đ	<O>+y	180
11	o	o	253	39	d	d	100
12	au	<O>+u, a	111	40	dh	<O>+z	189
13	h	<O>+h	250	41	n	n	116
14	m	<O>+m	181	42	p	p	112
15	ṃ	<S>+<O>+0	226	43	ph	<O>+p	185
16	ṃṛ	<S>+<O>+1	218	44	b	b	98
17	ṇl	<S>+<O>+2	219	45	bh	<O>+b	186
18	ṇv	<S>+<O>+3	220	46	y	m	109
19	ṇś	<S>+<O>+4	221	47	r	y	121
20	ṇs	<S>+<O>+5	222	48	l	r	114
22	k	k	107	49	l	l	108
21	ṁh	<S>+<O>+6	223	50	!	<O>+l	194
23	kh	<O>+k	251	51	!	<O>+L	241
24	g	g	103	52	v	v	118
25	gh	<O>+9	187	53	ś	<O>+s	167
26	ñ	<O>+g	169	54	ś	<O>+x	197
27	c	c	99	55	s	s	115
28	ch	<O>+c	141	56	h	h	104

は当該研究で重要な電子化テキスト等は未だユニコードベースで作成されるのはまれである。ある程度正確に作成された従来の電子化テキストをユニコードベースに変換することにより、計算できないエラーを招来する危険性がある。それゆえ、言語学解析の研究の現状を考慮して、我々の研究でもユニコードベースのフォントを主とした研究には切り替えていない。なお、Pali96フォントを含めたTrue Type フォント自体は従来のASCIIコードに割り当てられていると同時に、ユニコードにも割り当てられているので、将来必要に応じて解析システムをユニコード対応に切り替えることは可能である。

我々のインド・アーリヤ語古文献のデータベース (IAMDB) は、Web上に公開したデータと従来の形式で出版した書籍群から校正されているが、本論文では、電子化テキストと使用解説付き解析ツールに焦点を絞って議論する。

2. 電子化テキスト

当該研究の情報化に当たっては、信頼できるテキストをコンピュータ上に作成・保存・活用することが重要である。その作成に当たっては、フォントを簡単に入力できる簡易エディタが必須のツールである。その意味でも我々の簡易エディタは大変有意義である [1]。これまで、アルダ・マガダテキストに比べて、多くのパーリテキストが各種のコンピュータに入力されている。電子化テキストの標準化等については、テキスト形態が極めて多様であることから、あまり議論されていない。唯言えることは、これまで紙に印刷されたテキストが研究の基本にあり、これに基づき研究が蓄積されている。これまでの研究の継続性の観点から、電子化テキスト作成に当たっては、従来のテキストで採用されている形式に基づき、それに付加情報をつけたものが最適であろう。我々は、このような観点に立ち電子化テキストを作成することにする。このことにより、従来の研究によつて蓄積された知識の継承と同時に、テキスト入力・校正等の作業の効率化が図られる。

電子化テキストは単に作成されるだけではなく不十分であり、正確なデータが要求される。非常に残念なことに、日本国内にて作成された当該重要文献の電子化テキストが十分な検証のないままにWeb 上にアップロードされているために、その検証されたテキストが英国の Web サイトに掲載されている。正確なデータの作成に当たっては、言語学者の最新の知識と同時に情報技術者との共同研究が不可欠であり、極めて重要な研究過程であり、抜かすことはできない行程である。この事情は、山崎の評論にその事情が要約されているので以下に掲載しておく [www.sendai.ct.ac.jp/~ousaka 参照]。

「... 現在、メインの仕事は中期インド語で書かれた文献の語彙やバーダの索引作成、それに韻律解析である。索引という人はテキストを

打ち込み、索引作成用のプログラムを書いてしまえばプログラムが勝手に索引を作ってくれると思っているらしい。しかし、わたしがよく用いるパリテキストにはミスプリントもあれば、校訂者の語彙表記の不統一、リプリント時における改竄等、入力前に解決しなければならない種々様々な問題が横たわっている。ましてより難解なジャイナ教聖典においては言わずもがなである。このような問題には、研究者として自己の持つ最大限の能力を發揮して問題解決に当たるのが学者の良心であろう。したがって、われわれの作成した語彙索引はテキストそのものの単なる索引ではなく、この分野の最高権威であるK. R. ノーマン教授の校閲を仰いでおり、信頼性の高いものと自負している。そして、テキストに横たわる問題点を指摘し、コリゲンダ（正誤表）を学会誌にも発表している。……」

我々は、言語学者により作成された初期段階のテキストに対して、索引作成をしたり、または韻律解析等を施して、徹底したチェックを行っている（次節参照）。そのようにして極めて正確な電子化テキストを作成し、それを公表したり、又はそれに基づき各種索引を作成している。以下に主要なテキスト形式について記述する。

図1aはジャイナ教の5大主要テキストの一つである『ウッタラッジャヤー』のデーヴァナガリー原典テキストを示している。この形態では計算機処理できないので、我々のグループでローマ字化したものを図1bに示している。この形式の電子化テキストは、単語の索引作成にも用いるが、主に韻律解析用又は詩脚の正順と逆順索引作成用に使用する。このテキストの作成基準として、できるだけ（デーヴァナガリー）テキストに類似した形式を採用する。韻律解析に必要な最小限の情報のみをテキストに付加することとする。結果として、ローマ字化テキストを迅速に入力できると同時に、タイプエラーを容易に修正できる。このテキストは韻律解析と索引作成にも使用する。

このテキストは以下の特性を持つ。(1)特殊記号‘§’と章番号を先頭にもつ行が章のタイトルを示

す。この記号と章番号が付加された情報である；(2)散文行では、行の最後に必ず特殊認識記号//4,1//を持たなければならない。*4,1はこの行が4ページ目の第1番目の散文行であることを示している；(3)1詩偈は2行となる。1行目と2行目の最後に特殊記号‘/’と‘\’を付加して、これらが第1半詩偈、第2半詩偈であることを示す。2個の記号‘/’の間の数字は、詩偈の通し番号を示す。これらの付加した情報を基に、詩偈の部分のみが韻律解析され、また索引のライン情報（語彙の出現位置情報）が決定される。特殊記号‘\’が半詩偈を2つのパートに分ける記号である。

このローマ字化テキストだけでも研究上極めて有用であるが、さらに詩脚分解用のマーク()も付していることが、このテキストの有用性をさらに高めている。この分解マークを付与することは、当該分野の第1級の研究者でないと不可能であるが、我々の研究グループでは研究推進のためにこの貴重なデータを敢えてWeb上で公開している。このデータを研究者はよく活用しているようである。我々のデータを公表後は、当該研究会・学会等では、詩脚分割マークが付されたデータの活用が普通となっている。共同研究者である山崎が、これらのデータは我々のWebサイトからダウンロードしたものであること、を発表者に確認している。しかし、これらのデータを活用した研究でも、その使用を明言し謝辞を記載することがまれであり、その実態をなかなか把握できない。

原典テキストの形態として、図1の他に図2, 3のものもある。図2が、『ダンマパダ』で採用されているテキスト形式である。このテキストは韻文のみより構成されている。この電子化テキストは、もともとはノーマン先生によってNormynフォントを使用して入力されたものである。従来の出版されているテキストの形を忠実に再現しており、当然のことながら、計算機による韻律解析は最初から念頭にないので、図1bのような韻律解析又は詩脚の正順と逆順索引作成に必要な情報は付加されていない。また、詩偈がパート（=pāda、詩脚）に分割されたり、されていなかったりと統一のと

॥ परिशहज्जायणं द्वितीयं अध्ययनम् ॥
सुयं मे आपसे तेऽप्न भगवत्या पद्ममसामै। इह सत्तु चारोर्ष पर्तीशता
समर्णेण भगवत्या महावीरेण कार्यत्वेण पवरया, जे भिक्षु सोऽचा नच्चा
परीक्षाद् २० अग्राणपरीक्षाद् २१ दंसणपरीक्षाद् २२ ॥
परीक्षाद् सत्त्वे कार्यत्वेण पवरया।
तं मे उदाहरित्वामि आशुपुरुषं सुणेद्देः॥ १ ॥
१ दिग्मित्रापरिग्रहे देहे तारली निक्षु अतर्यं।
प छिद्रे प छिर्वापद प वृद्ध प व्यापद॥ २ ॥

एए परीक्षाद् सत्त्वे कार्यत्वेण पवरया।
जे भिक्षु स विद्वमेत्तजा उडु कण्ठं कण्ठद्वे॥ ३ ॥ त्वं बोमे॥
॥ परीक्षाद् समर्णं ॥ ४ ॥

図1a. ジャイナ教聖典『ウッタラッジャヤー』のデーヴァナガリーテキスト

॥ १ ॥ परिशहज्जायणं द्वितीयं अध्ययनम् /
सुयं me आपसे तेऽप्न भगवत्या पद्ममसामै। iha khalu bāvisam parisahā //4,1//
समर्णेण भगवत्या महावीरेण कार्यत्वेण पवरया, je bhikkhū soccā naccā //4,2//

parisahe 20 annāparisahe 21 dāmsanāparisahe 22 //5.8//
parisahānam pavibhatti kāsavenām paveiyā /
tām bhe udāharissāmi | āṇupuvvīm supēhi me //1//
l digimchāparigae dehe | tavassi bhikkhū thāmavāp |
na chinde na chindāvāc | na pae na payāvāe //2//

ee parisahā savve | kāsavenā niveiyā /
je bhikkhū na vihammejjā | puttho kenai kañhui //46// tti bemi /
/ parisahajjhāyanām samattām //2//

図1b. ジャイナ教聖典『ウッタラッジャヤー』のローマ字化テキスト

れていない表示になっている。我々のフォント系で書き換えると同時に計算機解析に必要な付加情報を作成する位置を丸印等でマークしている。計算機解析の目的に応じてテキストの各行の表示を統一する。例えば、韻律解析では、各半詩偈を1行で表示し、それに基づき解析を実行する。一方、索引等の作成のときは、各パーダを1行にして表示する。このことにより、従来の研究で採用されているように、各語彙の出現位置（ライン情報）を詩偈の通し番号とパーダ番号を使って表わすことが容易にできる。これらのテキスト形式の書き換え作業の一部（フォント書き換え等）はプログラムを作成することにより実行し、一部（章の記号‘§’の付加）は手作業でしている。

図3が、『ヴィナヤ』テキストで採用されているテキスト形式である。これは、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）経由で提供されたものであり、パリー聖典協会から出版されているテキスト形式を忠実に再現している。このテキストは3MBもの大量のデータ量である。このテキストについては、単語の正順索引作成が主たる計算機解析の目的である。それゆえ、図1bのような繁雑な付加情報を付加する作業を必要としない。付加情報としては、テキストのページを明示するマークを付けることだけでよい。これには、「§」をページを表わす情報として採用し、たとえば‘§218’が第218ページ目であることを示すようにする。又当然のことながら自前のプログラムを使用してNormynフォントから我々のフォントへの自動書き換え作業を実行する。

電子化テキストは、我々が作成したもの、海外の共同研究者が作成したもの等があり、そのテキストスタイルも上で述べたものだけでなく原典に応じて非常に多様である。目的に応じて、解析に適した形の付加情報を付加している[1]。しかし共通事項として言えることは、どの電子化テキ

- (7) subhānupassīñ viharantam indriyesh asaçplutam
bhojanamhi cāmattāññūñ kusitāñ hinaviriyam
tan ve pasahati māro vāto rukkham vā dubbalam.
- (19).bahum pī ce sahitam blāsamāno
na takkarō hoti naro pamatto
gopo va gāvo ganayāñ paresam
na bhāgavā sāmaññassa hoti.
- (20).appam pī ce sahitam bhāsamāno
dhammassa hoti anudhammacāri
rāgañ ca dosañ ca pahāya moham
sammappajāno suvimuttacitto
anupādiyāno idha vā huram vā
sa bhāgavā sāmaññassa hoti.
- (21).appamādo amatapadam pamādo maccuno padam.
appamattā na miyanti ye pamattā yathā matā.

- pāda a + pāda b
pāda c + pāda d
pāda e + pāda f
- pāda a
pāda b
pāda c
pāda d
- pāda a
pāda b
pāda c
pāda d
pāda e
pāda f

図2. 仏教聖典「ダンマバダ」の電子化テキスト

トでも、その著作権については細心の注意を払っている、ことである。著作権で保護されているものについては、その所有者に許可を貰い、通常は我々の研究にその使用が限定されており、他者に譲渡しないという条件が付いている。又、著作権の有効期限が切れているが、共同研究者が作成したものについては、その公開については共同研究者の意志を尊重している。このように、共同研究者及び海外の提携機関との信頼関係を損なわないよう、細心の注意を払っている。

この節の最後に電子ブックについて記述する。1997年頃にマッキントッシュを使って、ジャイナ教の5大聖典に関する5冊の電子ブックを作成したが、これにはアルダ・マガダ語文献に関する研究資料を網羅している。それに含まれる研究資料は、以下の通りである：ローマ字化された原典テキスト、韻律解析の解説と韻律解析結果、語彙の

§218 mahāvagga.
suppiyā upāsakassā nivesanārā ten' upasamkamī, upasamkamītā paññātē āsane nisidī saddhīm bhikkhusamghena.
"5" atha kho suppiyo upāsako yena bhagavā ten' upasamkamī, upasamkamītā bhagavantam abhivādetva ekamantam atthāsi. ekamantam thitān kho suppiyām upāsakam bhagavā etad avoca: kaham suppiyā 'ti. gilāñā bhagavā 'ti. tena hi āgacchatū 'ti. na bhagavā ussahatiti. tena hi parigaghetvāpī ānethā 'ti. atha kho suppiyo upāsako suppiyām upāsakam parigaghetvāpī ānete. tassa saha dassanena bhagavato tāvamahā vapo rūlho ahosi succchavi lomajāto.
"6" atha kho suppiyo upāsako suppiyā ca upāsikā acchariyām vata bho abbhutam vata bho tathāgatassa mahi-

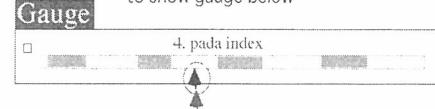
sannānam vā pasādāya. vigarahitvā dhammikatham katvā bhi-kkhū amantesi: santi bhikkhave manussā saddhā piṇḍāna. te hi attano pi marpāni paricattāni. na bhikkhave manussāmarpañi paribhūñitabbam. yo paribhūñjeyya, āpatti thullacevayassa. na ca bhikkhave appativedekhītā marpāni paribhūñitabbam. yo paribhūñjeyya, āpatti dukkā-tassā 'ti. "9" tena kho pana samayena rāñño hatthi maranti.

§219 mahāvagga.
manussā dubbhikkhe hatthimarpāñ paribhūñjanti.

図3. 仏教聖典『ヴィナヤ・ビタカ』の電子化テキスト



to show gauge below



This arrow shows which page of the book is now opened. By a drag of this arrow you can go to any page you will desire.

図4. 電子ブック『ウッタラッジャーヤー』

正順と逆順索引、詩脚の正順と逆順索引、詩脚索引プログラム、電子ブックには従来の本のスタイルをソフト的に実現している、ボイジャーのエキスバンドブックを使用している(図4参照)。図はローマ字化された原典テキストを示している。このスタイルの電子ブックは、違和感が無く大変使い勝手がよく、海外の研究者からも非常に好評であり、公表の初期の段階から使用されていた。一方、ウインドウズ系では残念ながら特殊フォントに起因するトラブルでうまく動作しない。ウインドウズでは同様の内容をPDF形式で代用している。

3. 使用解説付き解析ツール

我々の構築した言語処理システムの主ツールは、韻律解析ツールと索引作成ツールである。これらのツールは、中期インド・アリアンのアルダ・マガダ語、パーリ語と仏教混淆梵語に対して適用できる。各言語の解析では基準とするアルファベットが異なり、アルファベットの結合により語句の順番が異動するという厄介なことがある。特にパーリ語と仏教混淆梵語の語彙の索引では、このような異動を取り込んでいる。各言語では、4種類の索引(語彙の正順索引・語彙の逆順索引・詩脚の正順索引・詩脚の逆順索引)を作成できる。一方、韻律解析ソフトは各言語に対して1つである。その違いは、基準とするアルファベットが異なるだけである。解析プログラム等は使用解説書付きで既に公表されている[1]。本節では最初に韻律解析、次いで索引作成について述べる。

3.1 韵律解析

韻律解析において、パーリ韻文の特性を抽出するために、その韻文を中期インドアリアン語に基づいて研究することは勿論のこととして、古典サンスクリットに基づく研究も不可欠である。この研究は多数の複雑な詩偈の韻律解析が必要であることから、計算機解析に非常に適している。我々

表2. 中期インドアリアン聖典の韻律スキーム。左図(表2a):Akṣaracchandas, 右上図(表2b):Mātrāchandas, 右下図(表2c):Ganacchandas

韻律名	韻律図式	音節数又はモーラ数
Śloka	uuuu u---u uuuu u---u u---v uuuu uuuu u---u uuuu u---u uuuu u---u u---u uuuu uuuu u---u u---u ---u uuuu u---u	半詩偈の音節数 = 1 6
Modified Śloka	接頭辞を考慮に入れて、1 7 音節 を1 6 音節に変換(韻律図式は ショローカと同じ) 連続する2音節を1長音に変換 (韻律図式はショローカと同じ)	半詩偈の音節数 = 1 7
Trīṣṭubh	下記の3グループの総ての組み合わせ Opening (u---v; ---v; v---; v---), Break (v---; v---; v---; v---; v---; v---) Cadence (v---; v---)	バーダの音節数 = 1 1
Jagati	下記の3グループの総ての組み合わせ Opening (v---; ---v; v---; ---v; v---), Break (v---; v---; v---; v---) Cadence (v---; v---; v---; v---)	バーダの音節数 = 1 2
	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 = 2 2 ~ 2 5

は、パーソナルコンピュータを使って聖典の韻律を自動解析し、その結果を見やすい形に整理して、言語学者に提示することを目的とする。これにより、細心の注意を必要とする単調な作業が大幅に軽減されることになる。特に、中期インド・アリアン語の韻律は古典サンスクリットに比べて、必ずしも規則的とは言えず、ポエティック・ライセンスのとり扱いが難しいので、余計にこの作業の電算処理化の利点がある。

中期インド・アリアン聖典の韻律研究には、まず最初にその韻律スキーム、つまり半詩偈ごとの音節パターンを計算しなければならない(図5参照)。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(表1参照)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'm'を後に伴う短母音はいつも長母音として読まれる。(d): (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。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。なお、韻律表示に必要な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2に示すような中期インド・アリアン諸語の韻律(3大韻律, Akṣaracchandas(表2a),

Mātrāchandas(表2b), Ganacchandas(表2c))に基づき解析を行った。Akṣaracchandasは、音節数に

pariṣahāñampavibhāti kāśaveṇapaveiyā /	← 第1半詩偈
tamp bhe udāharissāmī āṇupuvūm sunēha me /1/	← 第2半詩偈
l digimchāparigae dehe tavassi bhikkhu thāmaṇam /	
na chinde na chindāvāe na pae na payāvāe /2/	

図5. 詩偈の構造

韻律名	韻律図式	音節数又はモーラ数
Vaitāliya	(Odd pāda) (Even pāda)	バーダのモーラ数 = 1 3 ~ 1 6
Aupacchanda-saka	w ---v u---v u w ---v u---v u	= 1 4 ~ 1 8
	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 = 2 6 ~ 3 4
韻律名	韻律図式	音節数又はモーラ数
Āryā	ww ww ww u www ww u www u u www u u	半詩偈のモーラ数 = 2 6, 2 7, 2 9, 3 0
Old Āryā	www www www u - , www www u www u u www u u	半詩偈のモーラ数 = 2 8 ~ 3 1

よって分類される。Sloka は半詩偈に 16 音節を含み、Tristubh と Jagati 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。Mātrāchandas 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 Aupacchandasaka と呼ぶ。Ganachandas 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2のスキームと比較して各詩偈の韻律種類を決定した(図8参照)。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題となる詩偈であるので、言語学者はこの韻律結果を活用して、研究をより効率的に推進できるであろう。またこの解析では、テキストのタイプミス等があったときに、韻律スキームに異常が現れてくる。よって、この解析はテキストの正確さをチェックする道具の役割も果たすことになる。

この韻律解析結果を参照して、典型的な Sloka は言語学者でなく情報科学者でさえも、半詩偈を詩脚に分解できる。Alternative Representation による表示が現れる半詩偈については、言語学者自身が慎重に詩脚に分解する。我々の研究では、言語学者自身が詩脚分解への確認を慎重に行って いる。この電子化テキストを活用することにより、

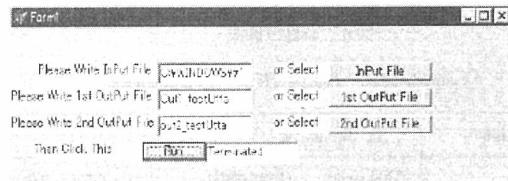


図6. 韻律解析用ダイアログウインドウ

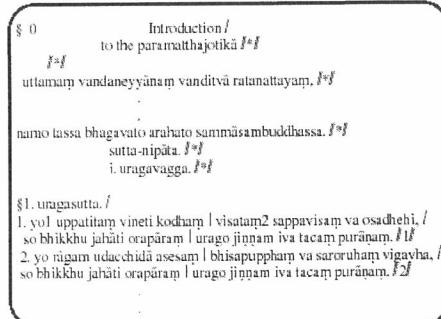


図7. 韻律解析用入力ファイル

詩脚索引を作成することができる。このようにして慎重に作成された正確なテキストが図1b等である。

韻律解析ツール操作は非常に簡単である。3つの言語に対する解析処理の中、パーリ語に対する処理を例にとって説明する。韻律解析ツールは適当な実行形式 jar ファイルをダブルクリックすることで起動し、図 6 に示すダイアログウインドウが現れる。1つの入力ファイルと2つの出力ファイルを指定すればよい。この場合、入力ファイルは例えば図 7 の形である。第1の出力ファイルには、図 8 のように韻律名・韻律パターンと解析した韻文がワンセットで表示されることになる。また第2のファイル（図 9）には、種々の韻律に対する解析を表形式に集計した結果を収録する。第1の出力は、各詩偈を詩脚に分割する際の基礎資料として利用できる。

3.2 語彙と詩脚索引作成ツール

これまで『ダンマバダ』、『ヴィナヤ』、『ディイガ・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ので、

図8. 一つ目の韻律解析結果出力ファイル

図9. 二つ目の韻律解析結果出力ファイル

これまで索引が作成されなかつたのが実状であった。我々は、言語学の研究者と情報科学の専門家との密接な共同研究により、短期間で大きなサイズのテキストの索引を完成させている。その他、我々のグループが成功した要因はもう1つある。次々と作成された索引を、出版・公表・配布できる中央学術研究所及びPTSと提携できたこと、も研究推進の上で大きな役割を果たしている。

これらの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』等については、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・多數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。最初に原典テキストと電子化テキストを比較して校正を実行する。それだけでは多くのミスを無くすことができないので、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す。誤植等のエラーは、索引又は詩脚索引上で、正常な語彙等に比べて特異な出現をするために、特定しやすい。この結果として、正確な電子化テキストを作成できると同時に、正確な索引を作成できる。現代の最高水準の言語学研究と情報処理を併用して、極めて正確な索引を作成している。

我々は上記の手順で、最初に『ダンマパダ』の4つの索引を求め1996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わらず、既に増刷されている。『ヴィナヤ』『ディーガ・ニカーヤ』『ジャータカ』『ヴィスッディマッガ』についてもその語彙の正順索引を順次PTSより出版した。

1	§Th/ 18	thera-gāthā /
2		dukanipāto. #*18.1
3		
4		
5	n' atthi koci bhavo nicco samkhārū vāpi sassatā, /	
6	uppañjanti ca te khandhā cavyanti aparāparam #121	
7	etam adinavānñ natvā bhaveñ amhi anathiko, /	
8	nissato sabbakāmehi, patto me āsavakkhayo #122	
9	itthān sudanñ āyasmā uttaro thero gāthāyo #*18.6	
10	abhaśithā 'ti. #*18.7	
11	na idam anayena jivitam, nābhāro hadayassa santiko, /	
12	āharatthitiko samussayo, tī disvāna carāmi esanam #123	
13	panko 'ti hi nam avedayum yāyan vandanapūjanā kulesu, /	
14	sukhumārya sallāpi durubhām, sakkāno kāpuresu dujjaho #124	
15	itthān sudanñ āyasmā piṇḍolabhbāradvāyo thero #*18.13	
16	gāthāyo abhaśithā 'ti. #*18.14	
17	makkato pañcavārayām i kuttikāyām pasakkaya /	
18	dvārena anupariyati ghatayanto muhūm muhūm #125	
19	tūṭha makkata mā dhāvi, na hi te tam yathā pure; /	
20	niggahito 'si paññāya, neto dūram gamissasi #126	
21	valliyō thero. #*18.19	
22	tinum me tilapatañāna gaṅgātire kuṭi katā, /	
23	chavaññito va me patto, parñukulañ ca cīvaram #127	
24	dvīnam antaravassānam ekā vācā me bhāsiñ; /	
25	tatiye antaravassāmhi tamokhandho padālido #128	
26	gaṅgātiriyo bhikkhu. #*18.24	
27	api ce hoti tevijo macchūyāñ anāsavo, /	
28	appaññāto 'ti narñ bālā avajānanti ajānatā #129	

図10. 語彙索引用入力ファイル

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』を中心にして”[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Webサイトに掲載している。パリ語文献に比べて、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』『ディーガ・ニカーヤ』『ジャータカ』『ヴィスッディマッガ』の語彙の逆順索引を作成し、Webサイトに掲載した。当初、マッキンツッシュ用の索引だけを掲載していたが、フランスの共同研究者(B. オグウェニン教授)からの要望により、この索引をウインドウズPCでも使用できるように拡張した。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. カズンズ教授から、すでに作成済みの『マハーニデッサ』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。この提案に基づき、当該テキストの索引を作成し、Webサイトに掲載した。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Webサイトへの掲載を見合わせている。

これらの索引作成に当たってのツール起動は、韻律解析の場合と同様、非常に簡単である。各言語毎に4種類の索引ツールがあるが、その構成は殆ど同じである。パリ語における正順索引ツールについて記述する。

パリ語における正順索引

ツールは、あらかじめ準備されたテキストファイル(図10参照)を処理して、図12の索引を作成する。このツールは2つのプログラムから構成されている。第1段階では、アルファベット順に単語索引を作成する。この段階では、図11に示す

1	'ti
2	§Th/ *18.7
3	§Th/ 124
4	§Th/ *18.14
5	§Th/ 129
6	'si
7	§Th/ 126
8	ajānatā
9	§Th/ 129
10	atthi
11	§Th/ 121
12	anathiko
13	§Th/ 122
14	anayena
15	§Th/ 123
16	anāsavo
17	§Th/ 129
18	anupariyeti
19	§Th/ 125
20	antaravasamhi
21	§Th/ 128
22	antaravasānam
23	§Th/ 128
24	aparāparam
25	§Th/ 121
26	api
27	§Th/ 129
28	appaññāto
29	§Th/ 129
30	abhaśithā
31	§Th/ *18.7
32	§Th/ *18.14

図11. 第1段階の語彙索引出力ファイル

1	'ti Th *18.7 124 *18.14 129
2	'si Th 126
3	ajānatā Th 129
4	atthi Th 121
5	anathiko Th 122
6	anayena Th 123
7	anāsavo Th 129
8	anupariyeti Th 125
9	antaravassamhi Th 128
10	antaravassānam Th 128
11	aparāparam Th 121
12	api Th 129
13	appaññāto Th 129
14	abhaśithā Th *18.7 *18.14

図12. 第2段階の語彙索引出力ファイル

ように、同一単語が何回も出現しているときは、見難い形になっている。第2段階では、この図11の結果を図12のように、より見やすい形にしている。これらの処理では、共に図13に示すようなダイアログウインドウを起動し、必要な入出力ファイルを指定し、Runボタンをクリックするだけで索引を作成する。第2段階の入力ファイルは、第1段階の出力ファイルを使用する。このように極めて、簡単な操作法であるので、今後当該分野の言語研究者の積極的な活用が期待される。

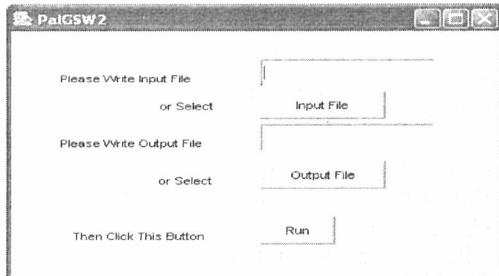


図13. 語彙索引作成用ダイアログウインドウ

4. 議論

我々はこれまで言語学研究者・情報科学研究者・国内外の研究者・機関との密接な共同研究の下に図14に示すような、中期インド・アリアン解析システム及びそのデータベースを構築してきた。それを活用して、この数十年間待望されていた『ダンマパダ』の語彙の正順・逆順索引と詩脚の正順・逆順索引、『ヴィナヤ・ピタカ』『ディーガ・ニカーヤ』等の語彙索引の作成に取りかかり、短期間でその作成に成功している。これら多数の第1次基礎資料は、今後長期間に亘って、学術的基礎研究等に対して多大な寄与することになり、中期インド・アリアン語の古文献に関する閉塞した研究状況を開拓し、インド学・仏教学研究を格段に発展させることができると予想される。

我々の開発してきたツールは、インターフェースを工夫してワープロ並に非常に使いやすい形になっているが、言語学者自身の使用自体はあまり

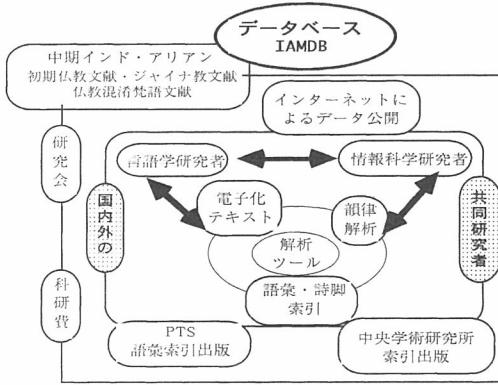


図14. データベースの構築組織

活発ではない。その為今後、我々の開発したツールがより一層使用されるような普及方法について、検討する必要があると思われる。

最後に、これらツール等の維持管理について述べておく。これまで議論してきたデータベースは所属機関のホームページ上に公開し、その管理・増補等は必要に応じて遂行している。このデータベースの当該研究における有用性から見て、将来的にその管理等ができなくなったときには、現時点では明快な答えはないのであるが、どのように処置するのかについて、その対策を考えておく必要があろうと思っている。一案として、オックスフォード大等のデジタルアーカイブに寄贈することも考えられる。

謝辞

この研究遂行に当たって適切な助言をしていたいた、山崎守一学長（宝仙学園短期大学）、宮尾正大教授（室蘭工大）に深く感謝いたします。本研究は2004年度日本学術振興会・科学研究費補助金、基盤（B）（科研費16320009）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, 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.
- 逢坂雄美「中期インド・アリアン古文献の計算機解析ツール—実行形式ファイルでの一般公開について—」『パーリ学仏教文化学』(2002) 第16号, pp. 109-116,
- 逢坂雄美「中期インド・アリアン古文献の計算機解析ツール—UNIXでのツール展開—」『パーリ学仏教文化学』(2003) 第17号, pp. 65-70,
- 逢坂雄美「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」『人文科学とコンピュータ』(2003-CH-59) pp. 55-60.
- [2] Y. Ousaka and M. Yamazaki, Index to the Visuddhimagga (Pali Text Society, Oxford 2004) 505pp.
- E. Fauré, B. Oguibénine, M. Yamazaki and Y. Ousaka, Mahāvastu-Avadāna, Vol. I: Word Index and Reverse Word Index, The Chuo Academic Research Institute (Tokyo, 2003) 282pp.
- Y. Ousaka and M. Yamazaki, Bhiksuni-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 Pali Canonical Texts (The Chuo Academic Research Institute, Tokyo 2000) 133pp.
- [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.