

インド・アーリヤ語古文献の研究に最適な統合的計算機 処理システム構築とその活用

逢坂 雄美

仙台電波工業高等専門学校
☎ 989-3124 仙台市青葉区上愛子北原

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、当該古文献を系統的に解析できるツール、①韻律解析と②語彙と詩脚の索引作成のためのツールを開発してきた。このツールは、開発者サイドに比重を置きすぎ、使いやすさが軽視されているために、言語学者が殆ど使いこなせない形式になっていた。この欠点解消のために、ファイルの入出力をダイアログ形式で簡単に指定できるように変更し、その後、実行（形式）プログラムを作成した。この結果、言語学者が情報学研究者の助けを借りずに、簡単に使える統合的計算機処理システムを作り上げることができた。このシステムは、マッキントッシュのみならずウインドウズ上でも稼働し、殆どの機種で使用可能であることから、当該研究の一層の進展に寄与すると期待される。

Construction of the Integrated Computer System for the Study of the Middle Indo-Aryan and its Application

Yumi OUSAKA

Sendai National College of Technology
Kami-Ayashi, Aobaku, Sendai 989-3124

In corroboration with the linguistic scholars and computer scientists, we have made the computer tools for the systematic analysis of the Middle Indo-Aryan (MIA) language, that is, metrical analysis tool and the tools for the productions of the word or pāda indexes. It was not taken into accounts how to easily run the tools by the users, regarding the file input and output, and set-up of the computer programs. In order to avoid these troublesome operation we adopted the dialog box for the manipulation of the file, and then made the execution files for these tools. As a result, we could construct such an integrated computer system that the linguist may easily manipulate for the study of the MIA without the help of the computer scientist. This system works well on not only Macintosh but Windows. This simple integrated system may be helpful for making more advance the study of the MIA.

1. 序論

当該古文献(B.C.5-A.D.10)には、サンスクリット語で書かれた膨大なウパニシャド哲学の文献を初めとし、プラークリット語で書かれたジャイナ教の文献、パーリ語で書かれた初期仏教聖典、それに仏教混淆梵語(Buddhist Hybrid Sanskrit)による大乘仏教経典が含まれている。これらの文献群は、世界文化の貴重な宝庫であり、現代文明の重要な源流の一つと見なされている。したがって、これらの文献の研究は仏教学を含むインド哲学の研究のみならず、比較言語や思想研究の分野にも不可欠となり、重要な意義を持つ。しかしながら、使用された言語が難解であること、異なった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、さらには量的に膨大であることもあり、ギリシャ・ラテン語文献になされたと同様な批判的研究が未だなされていないのが現状である。

アメリカ・日本・タイ等でテキストの電子化やCD-ROM出版等が盛んになされ始めた。言語学研究者自身は計算機を、論文作成・テキスト検索等に活用し始めた段階であり、これだけでは単なるワープロ機能を使用しているだけであると考えられる。これでは学術的基礎研究等に、現在のすぐれた計算機環境を十分に活用しているとはいえない。このように、言語学研究のための本格的な計算機ツール等について、国内外で殆ど研究されていない。

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、当該古文献を系統的に解

析できるツールを開発した[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、言語解析の3大要素、語彙と詩脚(=パーダ:音楽における1小節に相当する)・韻律・文法、を解析できる。①韻律解析は批判的校訂本の編纂に不可欠であり、②語彙の索引は翻訳に当たって非常な手助けとなり、③語彙の逆順索引は文法構造解析に必要となる。④詩脚の正順と逆順索引はテキストの正しい読みの確立と並行詩脚の検索に当って大変重要である(第1次基礎資料)。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。また、文法解析には、語彙の索引を入力ファイルとして活用する。各ツールは互いに関係しあっており、このような有機的で相互補完的な解析システムは、世界でも我々が唯一のものと思われる。

これらのツールを使用してこれまで、20冊ほどの基礎資料[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]。これらの研究により、言語学研究のための基礎資料作成に当たっては、計算機が相当の威力を発揮することが認識されるようになった。しかし、相変わらず国内外の研究者は、本を活用した従前通りの手法、必要事項を索引またはカードを繰ってチェックするという効率の悪い手法を採用している。この状況は我々の構築しようとしている言語学研究のための統合的な計算機処理システムにより、本質的に改善されると思われる。

我々はマックからウインドウズファイルへの変換等の目的で、ウインドウズで動作する簡易ワープロEDISANを開発した[3]。一方、言語学者自身もEDISANを使用して、これまで彼らがマック上に蓄積したデータをウ

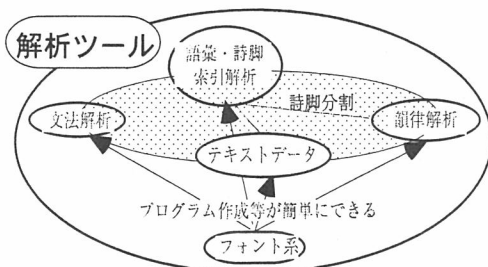


図1. 解析ツール

インドゥズファイルへ変換することに活用している。EDISANは特殊フォントも扱えるように特化しているが、比較的馴染みやすいコンパクトなワープロである。このことから、従来の解析ツールは情報学の研究者が操作しているが、韻律解析・索引作成プログラム等を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ときに、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。

本研究では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統合的計算機処理システムを構築することを目的とする。マック上では、殆どすべてのワープロが特殊フォントを支障なく処理できるので、上記①～④を処理できるシステムを構築する。一方、ウインドウ上では、特殊フォントを処理できるワープロがないので、図1に簡易ワープロEDISANを付加する。

このツールは、開発者サイドに比重を置きすぎ、使いやすさが軽視されているために、言語学者が使用するには抵抗感が強すぎた。このことは想像以上にツールの普及を妨げている。今後電卓並の使い勝手でツールを使用できるようになるならば、当該言語研究に対するIT効果が顕著に現れるようになると予測される。言語学研究者には、韻律解析ソフト・語彙と詩脚の索引作成ソフト・文法解析ソフトにおける入出力ファイルの処理、及びプログラムの起動・実行時の操作手順が面倒である。プログラムを実行形式に整理することにより、このような面倒な操作を簡単化でき、ワープロ並に使いやすいものとなるであろう。言語学者による機能チェックを考慮して研究を遂行する。

第2節では、言語研究の統合的計算機処理システムについて記述する。第3節では語

彙の順列索引作成ソフトの機能を例に取りその使用方法について述べ、第4節では韻律解析ソフトの機能等について議論する。第5節で問題点等について議論する。

2. 言語研究の統合的計算機処理システム

本研究で構築した言語処理システムは、図2に示すような索引作成ツールと韻律解析ツールから構成される。

各ツール作成には、統合的開発プログラムCodeWarriorのMPW Pascalを使用して



図2. マッキントッシュ用プログラム構成

いる。このシステムは、マッキントッシュとウィンドウズの両方で同等に動作し、異なるプラットフォーム上に同等な実行形式のファイルを作成できる。その実行形式ファイルは、単にこのファイルのアイコンをダブルクリックすることによって動作し、入出力ファイルの設定にはダイアログボックスを使用できるので、非常に使いやすくなる。また、このシステムのエディタは非常に優れており、通常のウィンドウズ上で動作するワープロと異なり、特殊フォントの処理も正常に行うことができる。以上のことに基づき、図2の言語処理システムをマッキントッシュ及びウィンドウズ上に構築した。

索引作成ツールは、中期インド・アリアン語のプラークリット語、パーリ語と仏教混淆梵語に対して適用できる。各言語のツールで

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。また、各順列索引作成には、次節で議論するように、3種類の実行形式ファイルを必要とする。一方、各逆順索引には4種類の実行形式ファイルを必要とする。この結果、マッキントッシュ・ウィンドウズ用と併せて、都合84（=3×(6+8)×2）種類の索引作成用の実行形式ファイルを必要とする。

一方、それに対して韻律解析ソフトは各言語に対して1つである。各言語間の違いは、基準とするアルファベットが異なるだけである。このツールでは、出力ファイルが2種類。これらのことについて第4節にて議論する。

Table 1. Pāli alphabet

Order	Character	Order	Character	Order	Character	Order	Character
1	a	14	ḥ	27	ṭh	40	m
2	ā	15	ṃ	28	ḍ	41	y
3	ā	16	k	29	ḍh	42	r
4	i	17	kh	30	ṇ	43	!
5	ī	18	g	31	t	44	l
6	ī	19	gh	32	th	45	v
7	u	20	ṅ	33	d	46	ś
8	ū	21	c	34	dh	47	ṣ
9	ū	22	ch	35	n	48	s
10	ṛ	23	j	36	p	49	h
11	ṛ	24	jh	37	ph		
12	e	25	ñ	38	b		
13	o	26	ṭ	39	bh		

3. 語彙の順列索引作成ツール

図2に示すように各言語毎に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ので、代表例として、パーリ語における語彙の正順索引ツールについて述べる。なお、我々のプログラムツールはすべて、Pali96フォントで作成されたテキストを対象としていることを、注意しておく。

パー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3参照）を処理して、図4の索引を作成する。このツールはファイルサイズの大きいものも処理できるように設計しており、3つのプログラムから構成されている。第1段階では、表1のアルファベット順に従い、かつ表2の特殊文字の組み合わせ処理を満たすように単語索引を作成する。この段階では、図5に示すように、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（例えば、ページ

Table 2. Special characters sequence

(1) ṃ + (k, kh, g and gh)	⇒	ṅ + (k, kh, g and gh)
(2) ṃ + (b, bh, p, ph and m)	⇒	m + (b, bh, p, ph and m)
(3) ṃ + (c, ch, j, jh and ñ)	⇒	ṅ + (c, ch, j, jh and ñ)
(4) ṃ + (d, dh, n, t and th)	⇒	n + (d, dh, n, t and th)

§1 namo tassa bhagavato arahato sammāsambuddhassa.

1. jātikotīśahashehi pamānarahitaṃ hitaṃ /1/
lokassa lokanāthena kataṃ yena mahesinā /
2. tassa pāde namassivā katvā dhammassa c' añjalim /2/
saṃghāñ ca patimānetvā sabbasammānabhājanam /
3. namassanādito | assa puññassa ratanattaye /3/
pavattassānubhāvena bhetvā2 sabbe upaddave /
4. taṃ taṃ kāraṇam āgamma desitāni jutimatā /4/
apañnakādīni purā jātakāni mahesinā /
5. yāni yesu ciraṃ satthā lokanittaraṇatthiko3 /5/
anante bodhisambhāre paripācesī4 nāyako /
6. tāni sabbāni ekajjhaṃ āropentehi saṃgahaṃ5 /6/
jātakam nāma saṃgitaṃ6 dhammasaṃgāhakehi7 yaṃ /
7. buddhavaṃsassa etassa icchantena ciratthitīṃ /7/
yācīto abhigantvāna therena atthadassinā /
8. asamsaṭṭhāvihārena sadā saddhivihārinā /8/
tath' eva buddhamittena santacittena viññunā /

図3. 入力ファイル例

amñāni I: 148-4

amsaṃ IV: 98-*16; 99-20 V: 320-9 -*16 -17

amsakūṭe I: 55-24 VI: 195-10; 392-21 -24; 425-20; 562-25

amsato VI: 557-3; 562-*13

amse I: 9-18; 158-9; 304-3 II: 13-14 -15; 117-20; 210-18

III: 179-1; 343-16 IV: 114-21 V: 132-15; 236-2 -24 VI:

21-10; 52-25 -25; 73-9; 524-29

amsen' VI: 343-5

amsena IV: 98-*16; 99-20; 137-26

amso II: 402-12

akamsu I: 2-19; 17-*16; 53-26; 81-31; 82-4; 115-1 -4;

122-8; 126-5; 127-19 -21 -25; 148-25; 178-16; 180-26;

188-16; 193-29; 198-15; 199-5; 200-10; 203-28; 207-1;

245-13 -15; 248-6; 250-24; 254-8; 259-11; 262-6; 264-24

-26; 266-19; 273-2; 283-2; 296-27; 315-17; 385-24; 399-

6 -25; 404-19 -29; 423-10; 428-29; 450-17; akamsū I:

181-8; 408-29 II: 437-3 III: 45-7; 215-21 IV: 318-23

VI: 105-30; 107-11

akakkasaṃ III: 282-*13 -15 V: 203-*27

図4. 索引の最終出力結果

'ham I 53-18 353-22

amñāni I 148-4

amsakūṭe I 55-24

amse I 9-18 158-9 304-3

akamsu I 2-19 17-*16 53-26 81-31 82-4 115-1 115-4

122-8 126-5 127-19 127-21 127-25 148-25 178-16 180-

26 188-16 193-29 198-15 199-5 200-10 203-28 207-1

245-13 245-15 248-6 250-24 254-8 259-11 262-6 264-

24 264-26 266-19 273-2 283-2 296-27 315-17 385-24

399-6 399-25 404-19 404-29 423-10 428-29 450-17

468-20

akamsū I 181-8 408-29

akataññussa I 319-15 322-*10

akataññussā I 322-12

akataññū I 295-19 295-22 297-24 298-*25 298-28 319-

17 319-20 319-20 322-20 322-21 326-15 465-21 465-24

465-24 467-8 467-14

akataññūjātaka I 377-13

akataññūjātakaṃ I 379-3

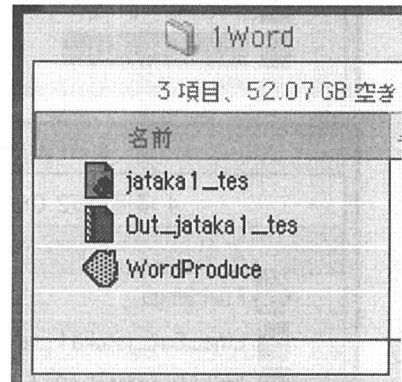
akataññūmittadūbhipuggalo I 322-7

図5. 第1段階終了後の索引結果

情報が 115-1 115-14 等のように、同一単語が同一ページに何回か出現したとき、その回数分だけページ番号が印字される。第2段階では、この図5の結果を図4のようにより見やすい形にしている。入力ファイルが大きくてボリュームごとに分割されているときは、第3段階の結合処理を実行する。図4はこの結果を示している。

図6に第1段階での操作手順を示してい

(a)



(b)

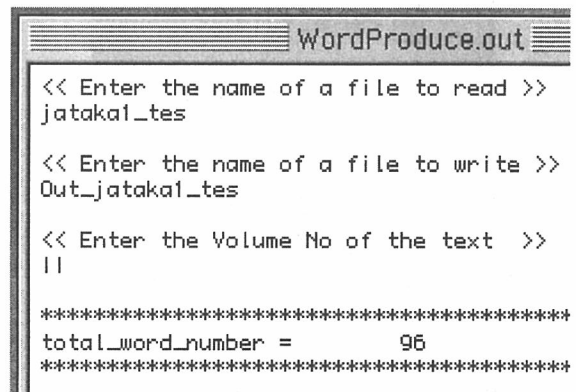


図6. 語彙索引作成の第1段階の処理

る。①最初に実行形式ファイルAwordProduceをダブルクリックする。②そうすると図6(b)のダイアログボックスの第1行目まで印字された画面が表示される。入力ファイル名を入力してリターンキーを押すと、③次に出力ファイル名の入力指示がでる。ファイル名を入力する。④次にボリューム番号の入力をする。⑤その後、プログラムが自動的にはしり、単語総数が96という表示が現れ、プログラ

イアログボックスが出現する。⑥第1段階で計算されたファイル(Out_jataka1_tes)が第2段階の入力ファイルとなる。その出力ファイル(Line_Out_jataka1_tes)名をまた入力すればよい。その結果、最初に空の出力ファイルが同じフォルダ内に自動的に作成され、その後処理結果が書き込まれる。⑦第2段階で処理されたファイルが最終段階の入力ファイルになる。これは1種のファイルマージ(結合)操作なので、2つの入力ファイル(Line_Out_jataka1_tes, Line_Out_jataka2_tes)と出力としての1つのファイル(Merge_12)を必要とする。

このように連続する3つの処理の結果、入力テキスト(図3)より最終結果(図4)が得られる。以上のように、殆どワープロ操作と同じ感覚で索引を得ることができることになる。これにより言語学者自身が抵抗感なくこのツールを使用できると思われる。

他の語彙の逆順索引・詩脚の順列索引及び逆順索引ツールも基本的には同じように動作する。

4. 韻律解析ツール

次に韻律解析ツールの処理について、図8に基づき説明する。ここでもまた、パリー語の処理を例に取り説明する。実行形式ファイルはAMetrePaliであり、前と同じくそのアイコンを2回クリックすることによりその動作が始まる。その下の図に示すように、索引処理と同様な操作により簡単に最終結果が得られることが分かるであろう。この場合、入力ファイルはIsn4(図9参照)であり、出力ファイルは2種類必要であり、(Out1_sn4, Out2_sn4)を使用している。最初の出力ファイルは図10の形式である。図9において、行末が/*//の場合散文を表し、/又は//1//等の場合は韻文を表す。韻律解析では、韻文に対してだけ解析処理をするので、図10のような結果を表示することになる。韻律名・韻律パ

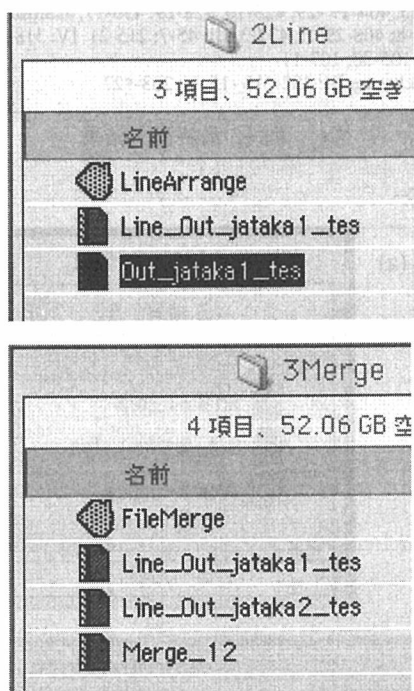


図7. 第2, 3段階の処理

ムが終了する。作成された索引は、図6(a)のフォルダに自動的に格納される。このファイルをダブルクリックすると結果をみることができる。

第2段階、3段階の処理用のフォルダ情報を図7に示す。第2, 3段階では、実行形式ファイルはそれぞれ、LineArrange, FileMergeであり、第1段階と同じくそれぞれのアイコンをダブルクリックすることにより動作を開始し、対応するファイル入出力用ダ

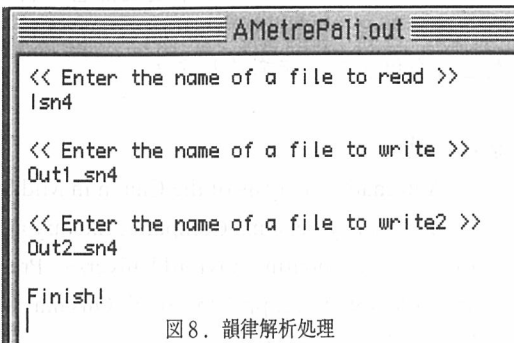


図 8. 韻律解析処理

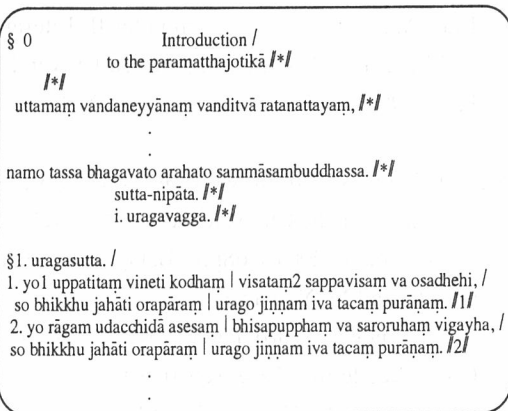


図 9. 韻律解析用入力

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。この解析の詳しい内容については、論文 [1] 参照のこと。また第 2 の出力結果を図 11 に示す。この結果は、種々の韻律に対する解析結果を表形式に集計した

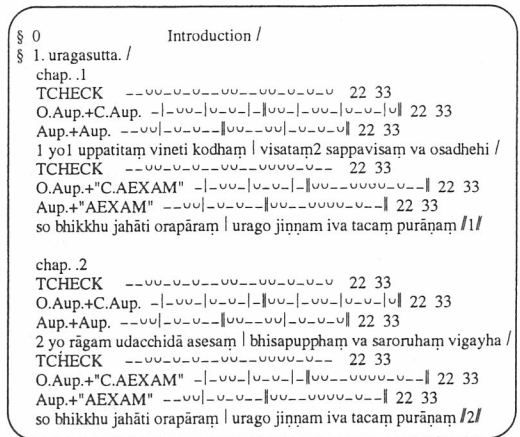


図 10. 韻律解析結果その 1

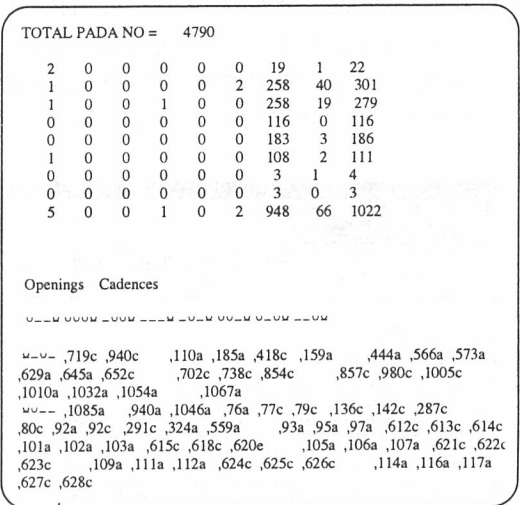


図 11. 韻律処理結果その 2

ものである。この意味についてもまた、論文 [1] 参照のこと。

同じような韻律解析用のツールが⁵、ブライクリット語及び仏教混交梵語用に準備されている。

なお、図 10 の解析結果は、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。

5. 討論

第 3 節、4 節で得られた計算結果はそのファイル形式のままでは利用しづらいので、適当なワープロにデータを移して研究に使用することとなる。マッキントッシュの場合、

殆どの市販されているワープロはその入力を含めて特殊フォントを不自由なく扱うことができる。

一方、ウインドウズ環境では、そのOS自体の制約から、特殊フォントの簡単なキーボード入力法さえない。それ故、ウインドウズシステムは特殊フォントを使わざるを得ない中期インド・アリアン語の研究には非常に不便である。その解消に向け、我々はこれまでウインドウズ環境下で動作する特殊フォント用の簡易エディタEDISAN(図1 2 参照)を開発した。EDISANはWordPadをベースに開発したので、基本的には他のワープロと同様に使い勝手がよい。ウインドウズ系では、図2で述べた言語研究の統合的環境にこのEDISANを付加することにより、マッキントッシュ系

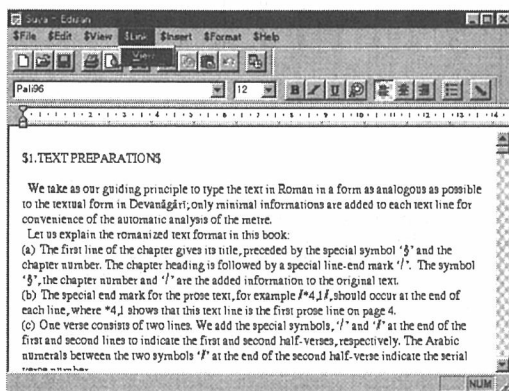


図1 2. 簡易エディタEDISAN3.0

と同等な統合環境が構築される。

言語学研究には原典と索引だけではなく、常時その他何冊もの文献を参照しながら、総合的に判断して研究(批判的校訂本作成等)を遂行する。この種の研究において、参照文献データをパソコンに取り込む必要が生じる。OCRでの読み込みが正確であれば、この作業が効率よく進むが、残念ながら現時点でもあまり読み込み精度がよいとはいえない。今後この読み込み精度が上昇すれば、ますます研究環境がよくなると思われる。

文法解析ツールは残念ながら索引作成と異なり、実行形式のファイルを作る段階まで達していない。今後このことについては研究を進める予定である。

最後に、本研究で作成した実行形式ファイル等は、今後そのマニュアルが完成次第公開・提供する予定である。

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、山崎守一教授(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。この研究は2001年度文部省・科学研究費補助金、基盤(B)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] Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer, Literary and Linguistic Computing (Oxford University Press 1994) Vol.9, No.2, pp.125-36, Y. Ousaka, M. Yamazaki and M. Miyao
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, Y. Ousaka and M. Yamazaki
- [2] Prātimokṣasūtram: word index and reverse word index (The Chuo Academic Research Institute, Tokyo 2001), 68pp., B. Oguibenine, Y. Ousaka and M. Yamazaki
A word index to early Pāli canonical texts (The Chuo Academic Research Institute, Tokyo 2000) 133pp., M. Yamazaki and Y. Ousaka
A pāda index and reverse pāda index to early Pāli Canonical Texts (Kosei Publishing, Tokyo 2000) 571p., M. Yamazaki and Y. Ousaka
- [3] ウインドウズ用特殊サンسكريットフォントのためのエディタ作成、『人文科学とコンピュータ』研究会、(2000.7.21) pp.23-29, 逢坂雄美, 鈴木享