

中国語とパソコン
—その古くて新しい関係—

内田慶市

中国語とパソコン

—その古くて新しい関係—

内田 慶市

1. コンピュータと漢字

1.1. ビットとバイト

コンピュータの基本的な処理は、1か0かによって行われる。つまり2進法で情報は表されるが、2進法の1桁を「ビット」と言い、nビットでは2のn乗個の情報が表されることになる。一般には8ビットを一つの単位とし、これを1「バイト」と呼んでいる。つまり、1バイトでは2の8乗=256個の情報が表わされる。

1.2. 「キャラクターコード」

ただ、この8桁からなる0と1の数の並びは読み書きに不便であり、実際には16進法で表記される。16進法では16種類の数字が必要になるが、0から9までは10進法の数字を使い、10以降はアルファベットのAからFまでを使用する。これは、すなわち4ビットの表わす情報量に等しい。以下の通りである。

2進法	16進法	2進法	16進法
0000	0	1000	8
0001	1	1001	9
0010	2	1010	A
0011	3	1011	B
0100	4	1100	C
0101	5	1101	D
0110	6	1110	E
0111	7	1111	F

また、8ビットを一列に書いたのでは見にくいいため、上位4ビットと下位4ビットにわけて表記するが、これによって表わされる256の情報をアルファベット、数字、カナに対応させたものを「キャラクターコード」と呼ぶ。(図1参照)

1.3. コンピュータと中国との古い関係—易とコンピュータ

ところで、いくらかでも中国や中国語に関心のある人ならば、このコンピュータの原理はどこかでお目にかかったことがあるように感ずるかも知れない。まさにその通りであり、コンピュータの原理は、陰陽二元論に基づく「易」(『易経』)のそれに通ずるものがある。このことは、藪内清氏(『中国の数学』1974)によれば、2進法の原理の考案者であるドイツの哲学者ライプニッツが夙に指摘していることではあるが、前近代的・非科学的と言われる「占い」と近代的・科学的な「コンピュータ」が、実は同じ原理に依っているというのはまことに興味深いものがある。

上位4ビット

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		DE	0	@	P	p						一	タ	ミ	三	×
1	S _H	D ₁	!	I	A	Q	a	q			。	ア	チ	ム	三	円
2	S _X	D ₂	"	2	B	R	b	r			フ	イ	ツ	メ	三	年
3	E _X	D ₃	#	3	C	S	c	s			」	ウ	テ	モ	三	月
4	E _T	D ₄	\$	4	D	T	d	t			、	エ	ト	ヤ	三	日
5	E _Q	N _K	%	5	E	U	e	u			・	オ	ナ	ユ	三	時
6	A _K	S _N	&	6	F	V	f	v				ラ	カ	ニ	三	分
7	B _L	E _B	/	7	G	W	g	w				ア	キ	ヌ	三	秒
8	B _S	C _N	(8	H	X	h	x				イ	ク	ネ	三	
9	H _T	E _M)	9	I	Y	i	y				ウ	ケ	ノ	三	
A	L _F	S _B	*	:	J	Z	j	z				エ	コ	ハ	三	
B	H _M	E _C	+	:	K	[k					オ	サ	ヒ	三	
C	C _L	→	,	<	L	¥	l	!				オ	ヤ	シ	三	
D	C _R	←	-	=	M]	m					ユ	ス	ヘ	三	
E	S _O	↑	.	>	N	^	n	~				ヨ	セ	ソ	三	
F	S _I	↓	/	?	O	_	o	~				ヨ	ツ	マ	三	

(図1)

坤 ☷ (地)	艮 ☶ (山)	坎 ☵ (水)	巽 ☴ (風)	震 ☳ (雷)	離 ☲ (火)	兌 ☱ (沢)	乾 ☰ (天)	上卦	下卦
☷	☶	☵	☴	☳	☲	☱	☰	乾(天)	乾(天)
135	241	94	121	31	159	93	39	乾为天	乾为天
☷	☶	☵	☴	☳	☲	☱	☰	天沢履	兌(沢)
194	81	220	226	180	59	200	128	天沢履	兌(沢)
☷	☶	☵	☴	☳	☲	☱	☰	天火同人	離(火)
44	214	241	52	187	268	144	151	天火同人	離(火)
☷	☶	☵	☴	☳	☲	☱	☰	天雷无妄	震(雷)
228	246	76	89	159	207	179	235	天雷无妄	震(雷)
☷	☶	☵	☴	☳	☲	☱	☰	天風姤	巽(風)
123	166	136	201	17	132	254	107	天風姤	巽(風)
☷	☶	☵	☴	☳	☲	☱	☰	天水訟	坎(水)
107	87	260	214	74	249	129	100	天水訟	坎(水)
☷	☶	☵	☴	☳	☲	☱	☰	天山遯	艮(山)
165	165	67	173	233	194	7	24	天山遯	艮(山)
☷	☶	☵	☴	☳	☲	☱	☰	天地否	坤(地)
64	221	114	201	172	38	115	144	天地否	坤(地)

(図2) 本田清「易」(朝日新聞社)より

兌 ☱	艮 ☶	離 ☲	坎 ☵	巽 ☴	震 ☳	坤 ☷	乾 ☰	
沢	山	火	水	風	雷	地	天	自然
少女	少男	中女	中男	長女	長男	母	父	人間
説	止	麗	陥	入	劬	順	健	風性
羊	狗	雉	豕	雞	亀	牛	馬	動物
口	手	目	耳	股	足	腹	首	身体
西	東北	南	北	東南	東	西南	西北	方角

(図3)

「易」の中心をなすものは「卦」であるが、「卦」は6本の棒からなっている。その棒には一と--の2種類があり、それぞれ「陽爻（剛爻）」「陰爻（柔爻）」と言い、相反する属性を表わす。その組み合わせは全部で64通りになり、この「六十四卦」によって世界の変化を表現するというわけである。（図2参照）

また、「六爻」からなる「卦」は実は「三爻」ずつのものが上下に重なってでき上がったもので、基本は「三爻」の組み合わせであり、その組み合わせは8通りで、これをいわゆる「八卦」と呼ぶのである。（図3参照）

こうしてみると、これはまさしく2進法の世界そのものであり、先のコンピュータの用語を使えば「易」の基本は「3ビット（=八卦）」であり、「上位3ビット（=上卦）」と「下位3ビット（=下卦）」で森羅万象を表現していると言うことになるであろう。

コンピュータ向きでないと思われがちな中国あるいは中国語の世界が、実は最もコンピュータに近いところにいたということは記憶に止めておいてもいいだろう。

1.4. 1バイト系文字と2バイト系文字

さて図1で示されたように、アルファベット、数字、カナは1バイトコードで表わされることから、「1バイト系文字」と呼ばれる。（Alphabet, Numeric, Kanaの頭文字をとって「ANK文字」とも言う）

一方、日本の漢字や、中国の漢字は256の情報では表示しきれない。元々コンピュータが「漢字圏」の言語を使用する人々によって考えられたものでないこととはいえ、私たちは、この点ですでに大きなハンディを背負わされているわけであるが、そこで、2バイトを使って表現することになる。2バイトとは16ビットだから、2の16乗=256×256=65536となり、これでようやく漢字の表示がされることになる。日本語や中国語が「2バイト系」と言われるのはこのようなことによる。ただし、異体字などを考えるとこれでも足りないという場合がおり、最近では「3バイト」あるいは「4バイト」を主張する人も出てきている。

1.5. 日本製パソコン（PC98シリーズ）での中国漢字の表示の仕方

私たちが使うパソコンは普通は日本製・日本仕様のものであり、これにはJIS漢字がROMとして搭載されている（もちろん、これはPC98シリーズを念頭においたものであり、英語仕様のIBMやその互換機、さらにDOS/VやMacではその限りではない）。これだけでも、1バイト系言語の処理に比べればコンピュータにとっては大きな負担になるのであるが、その上更に中国の漢字を表示させるようにするのは実に大変なことである。

今、後に詳しくその機能等には触れる「文華13000」を例にとり、日本製パソコンによる中国漢字表示の仕組みについて簡単に述べておく。

(a)アプリケーションの中での処理

アプリケーション（＝「文華13000」）の中では、中国漢字データを別に用意し（いわゆる「簡体字・簡体字ROMボード」）、グラフィック処理によってすべての中国漢字をビットイメージ、つまり「絵」として表示させる。

なお、入力コードは次の図4の表のようになっている。GB1区からGB94区（GB＝「中華人民共和国国家標準信息交換用漢字編碼字符集」）までが中国漢字コード領域にしているが、GB12区～15区、および88区～94区は未定義であり、拡張外字領域となっている。（図4参照）

(b)テキスト画面上での処理—「フロント文華」の場合

アプリケーションの中ではグラフィック処理でスピードが遅いことを除けば、さほど問題はないが、中国語部分をアプリケーションから切り放してFEPとしてテキスト画面上（＝DOSコマンド上）で中国漢字を表示（これが「フロント文華」である）させようとすれば、

B1 ↓	B2 →	00	3F	40	7E	7F	80	FC	FD	FF
		00 ⋮ 80								
		81 ⋮ 9F		JIS			JIS			
		A0 ⋮ BF		GB1区~64区			GB1区~64区			
		C0 ⋮ DF		声調つき拼音記号			声調つき拼音記号			
		E0 ⋮ EA		JIS			JIS			
		EB EC		PC98 外字			PC98 外字			
		ED EE		拡張漢字			拡張漢字			
		EF								
		F0 ⋮ FE		GB65区~94区			GB65区~94区			
		FF								

(図4) 「文华13000」コード表

グラフィック処理はできないため「外字」の機能を利用しなければならない。ところがMS-DOS (NEC) では「外字」は188字しか利用できず、中国漢字をすべて表示させることは当然できないわけである。そこで「フロント文華」では、足りない部分は JIS 漢字で代用させるという方法を取っている。しかし、これはあくまでも画面表示上の処理であって、入力されるコードや印刷される漢字はあくまでも中国漢字 (ROM ボードによる) そのものである。

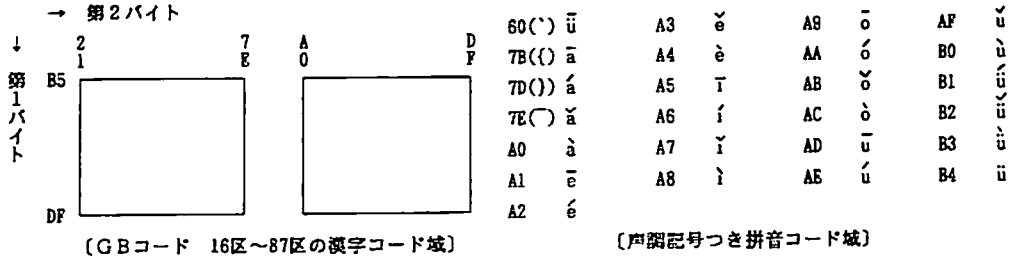
さて入力コードであるが、画面表示と同様に日本製パソコンには中国漢字用コード域を持たないことにより、またまた JIS のお世話になることになる。すなわち、JIS コード (厳密にはシフト JIS) の一部に中国漢字を割り当てるわけで、「フロント文華」では半角カタカナの領域 (B5から DF まで) にこれを当て、「半角カタカナ+1 バイト」で漢字1字を表現している。例えば、次の通りである。

GB 区位	→	JIS	=	内部的文字	→	中国漢字
1601		B521		オ!		啊
1602		B522		オ"		阿
1603		B523		オ#		埃
1604		B524		オ\$		挨
⋮		⋮		⋮		⋮
3667		C153		チS		你
⋮		⋮		⋮		⋮

この例を見ればわかるように、「阿」「埃」「挨」などは日本の漢字にもあるのであるが、内

部コードは違っているのであり（「挨」は JIS で3027、シフト JIS で88A5）、画面表示の時だけ日本のもので代用させ、実際の印刷などの場合には正しい中国漢字を使用するといった実に手間のかかる、涙ぐましい処理がなされていることがわかる。「中国漢字、その本性見たり、半角カナ」である。

なお、声調付ピンインも漢字と同じようにカタカナや記号のコードに割り当てられている。（図5参照）



(図5)「フロント文華」コード表

以上のように、「漢字」をコンピュータで処理していくのは実に骨の折れることであることが理解できると思うが、それでも多くの Developer 達の努力によって、ようやく1バイト系言語と同じくらいのレベルの処理が望めるようになってきている。

2. 現在入手できる中国語ソフト

現在日本で入手可能な、パソコン上で動く中国語ソフトはそれほど多くはないが、代表的なもの（とくに筆者の使用経験のあるもの）について概観しておく。

2.1. 「文華13000」(プロビット)

PC-98シリーズ用中国語+日本語ワープロソフトとして、1985年にまず「文華」で登場し、その後、さらに「文華13000」「パーソナル文華」となる。

初代の「文華」は日本の JIS 第2水準のROMの部分中国漢字（簡体字）に入れ替えるというハードの変更を伴うものであった。こうすることによって、すべてのアプリケーションの中だけでなく、DOS上、つまりテキスト画面上でも中国漢字は問題なく表示され、当然DOSコマンドも利用できるし、その上スピードの面でも申し分ないものであった。ただ、この場合、使える文字種は JIS 第1水準2965字と中国の簡体字第1水準のみ（厳密に言えば第1水準簡体字3755字+第2水準簡体字の一部の659字+非漢字684字）というのがネックとなった。これではいかにも「少なすぎる」わけで、必要なものは「外字」として作字をしなければならず、しかも「外字」にも「188」字という制限があり、実用という点からはやはりいささか不便であった。

「文華13000」では、この文字種の量の問題は大きく改善された。「文華13000」の扱える文字種は JIS 第1、第2水準漢字6349字+非漢字645字に加えて、中国の第1、第2水準簡体字6763字+非漢字816字であり、まさにその名の通り「13000」の日中両国の漢字が表示できるようになったのである。なお、中国漢字のデータは拡張ボードに漢字ROMボードで供給されるため、拡張ボードのない機種（たとえばノート型）では使用できないという制限はあるが、初代「文華」ほどではないにしても、スピードの面でもさほど問題はなかった。（漢字データをソフトで持つのに比べれば、ROMの方がやはり速い）

また、「文華13000」では、「簡体字ボード」に加えて「繁体字ボード」も取り付けが可能になり、その結果「簡体字」「繁体字」両方の表示、切り替えも自由に行われるようになった。なお、印刷の面でも、プリンタの制限はあるものの48ドットフォントが用意されて48ドット印字もできるようになった。

最近人気のノート型、ラップトップ型用のものとして「パーソナル文華」があるが、これは漢字 ROM の取り付けができないことにより様々な制限があり、とりわけ日本語との混在ができない点(コード入力を使えば可能ではあるが、日本語の FEP が使えないので実用にならない)、また「フロント文華」も使えないという点で問題が残る。(最近「パーソナル文華」は、出荷を取り止めているようである)

「文華」シリーズの中国語入力方法は基本的にはピンイン入力であるが、そのほかピンインあいまい入力、部首別/G B 区位別/G B 区位コード入力、電碼(中国の電報コード)入力、さらには日本語読み入力もサポートしている。また、「文華13000」からは、声調付ピンインも表示でき、中国語のテキスト作りなどには便利になっている。ただ、この声調付ピンインでの難点を言えば、大文字母音の声調が表示できないことである。(たとえば、Ē, Ā, Ēi は ē, ā, éi としかならない)

さて、「文華」シリーズで私たちにとって最も有益なのは、先にすでに述べたように、中国語漢字変換部分(FEP)を切り離して使えるという点である。CONFIG.SYS に次の2行を加えることによって、他のアプリケーション(ワープロソフトは別)や DOS 上での中国語漢字表示が可能になるわけである。

```
DEVICE=FWENSUB.SYS
```

```
DEVICE=FWEN.SYS [パラメータスイッチ]
```

さらに日本語の FEP を加えるには、中国語変換デバイスの中に、日本語変換ドライバーを書き込めばよい。たとえば、次の通りである。

```
DEVICE=FWENSUB.SYS
```

```
DEVICE=WXK.SYS [パラメータスイッチ]
```

```
DEVICE=WX2.SYS [パラメータスイッチ]
```

```
DEVICE=FWEN.SYS [パラメータスイッチ]
```

これで、データベースソフトや、エディタ上でも日本語、中国語両方の表示が可能になるのであるが、日本語モードでは中国漢字は文字化けが起こるし、中国語モードでは日本語の半角カタカナが中国漢字に化けるなどの点はいたし方ないことである。(理由は先に述べた)また、エディタなどでは、中国語モードの際に、カーソルキーが殺されるために、スクロールなどは Ctrl キーを使った「ダイヤモンドカーソル」を利用しなければならない。

なお、データベースソフトでは、「”」「,」などを項目の切れ目として使用している場合が多く、中国漢字のうち、これらの文字をその構成コードとして持っている時には、正しく処理できないということもあるのであるが、この点は、プロビットでそのコードをはずした改良版を試供してくれる。たとえば、「李」「文」はそれぞれ内部コードは「BF22」「C72C」となっており、第2バイト目に「”」「,」を含んでいるが、試作改良版では「BF64」「C7BF」と改められている。ただ、この場合、「文華13000」のコードとの調整が済んでおらず、コンバートの際にずれが生ずることは承知しておかねばならない。

2.1.1 「フロント文華」と Micro-OCP を使った例

ここで、「フロント文華」とデータ処理ソフトを組み合わせて行った中国語処理の1例を示しておくこととする。

データ処理ソフトとして使用したものは、「Micro-OCP」という「文章解析プログラム」（沖田電子技研）である。「OCP」は「Oxford Concordance Program」の略称で、オックスフォード大学で開発されたプログラムであり、英国の全ての大学や、日本でも東京大学大型計算機センター、東京都立大学、北海道大学、東京女子大学などでも設置されており、これをパソコン用に改良したのが「Micro-OCP」である。

「OCP」には、次の三つの基本的な機能がある。

(1) 単語の一覧表と出現頻度表の作成 (WORD LIST)

(2) 索引作成 (INDEX)

(3) 用語索引作成 (CONCORDANCE)

さらに、この応用で語彙の分類や、統計、語の出現順序、パターン・マッチング、ひいては著者推定、真偽問題、作品の年代の特定といったことも可能になるという。

また、「OCP」の処理命令は次の五つの定義部からなっている。

• INPUT：テキストデータの構成の指定

• WORDS：テキストデータの記述に使用されている文字や記号の指定

• ACTION：OCP の解析方法の指定

• FORMAT：OCP 実行結果の出力形式の指定

• GO：OCP 文章解析の実行開始

今詳しい説明はしないが、要するに文章データを文献の通り（行数、文字数、ページ数など）入力してやり、INPUT と WORDS で区切り文字などを指定し、ACTION で目的の処理を指定し、さらに FORMAT で出力結果を指定するということである。これだけの簡単な操作によって、たとえば後ろの資料(A)に挙げたような結果が得られるのである。A.1は文章データ、A.2は INDEX、A.3は PICK WORDS で指定された語彙の CONCORDANCE、A.4は PHRASE や共出語 (COLLOCATE) の抽出である。

中国語文献の処理の場合には、まず「文華13000」で、原文に忠実に入力していく。

この時、単語の切れ目として「全角のスペース」或いは任意の全角文字（記号）を用いる。「半角スペース」でもいいのであるが、「文華」の中国漢字コードがすべて半角コードの組み合わせであることから、後のテキストファイルコンバートの際の誤認識を避けるためである。また、単語の切れ目は、研究の目的、さらには各人の「単語認定の基準」によって決まってくる。これは、英語などの処理とは異なる点である。日本語でも同様の問題が生じてくるであろう。句読点は区切り文字として認識されるから、そのまま入力する。改行も原文通りに行うが、1行が長すぎて強制改行されてしまう場合などの時には、改行を示す文字（記号）を決めておき、それを入力してもよい。改行はいつも「リターン」でなくてもかまわないのである。なお、書名や編・章名などをあらかじめ定義しておく。

次に、でき上がったデータを、「フロント文華」ファイルにコンバートする。そのあと、「フロント文華」を組み込んだ状態で「Micro-OCP」で目的の処理を行えばいいわけである。

「用語索引 (CONCORDANCE)」でも、単語の抽出には、中国語モードで行えば中国漢字が表示されるので何ら支障はないはずである。ただ、「Micro-OCP」は日本語仕様であるため、WORDS の指定で若干の工夫が必要である。特に半角カナコードを中国漢字がその構成コードとしてもっているのを、それを1バイト文字でない旨の記述をしてやればよい。このようにして処理された結果を、再度「文華13000」ファイルにコンバートすれば完了である。

(実は「文華13000」にコンバートせずとも、フロント文華が組み込まれていれば、フロント文華用のファイルのままで、直接 OCP やエディタで印刷することもできるし、その方が文書量の制限はなくなる)

後ろの資料(B)は、私が以前行った、『華語拼字妙法』(1911)という「官話」資料の OCP による処理の例である。

B.1は入力した原文の一部、〈T〉は「書名」、〈K〉は「課」、〈P〉は「ページ数」、〈L〉は「行数カウント」の始まりの定義である。(改行はリターンを使っているが、印刷の関係で資料では1行に収まらなかった)また、「=」は改行しているが、次の行に来る語と「1語」であることを示している。

B.2は、コマンドである。「用語索引」のコマンドでは、抽出する語(田地、晓得、成功、让)が文字化けしているが、これは今回はフロント文華を組み込まずに印刷したためであり、処理上何ら問題はない。

B.3は、「用語索引 (CONCORDANCE)」のサンプルである。

B.4は、「語彙索引 (INDEX) (初稿)」の一部である。

なお、中国語を処理する場合、用語索引などで、OUTPUT される各行の両端が文字化けしたり(これは、各行の文字数がきちり偶数であれば、こういうことにはならないのであるが、奇数の時には、2バイトが分かれてしまうことによる。つまり、半角コードが残る)、単語の切れ目に使った「全角スペース」は取り除きたいということがあがるが、これらは、MS-DOS の TOOL である SED などを使えば奇麗に整形がなされる。(これについては大東文化大学の中村浩一氏の一連の発言が、パソコン通信ネット Nifty の FMCN フォーラム11番会議室にみられる)

また、フロント文華と dBASE を組み合わせた中国語データベースソフトも市販されてきており、たとえば「電腦司書」(日中産業通信)がある。

2.2. 「KOA-TechnoMate2中国語 (TM中国語)」(高電社+プロビット)

これは、これまで多言語ワープロソフトとして定評のあった「KOA-TechnoMate2」の中国語バージョンであり、中国語部分は上の「文華13000」を移入したものである。このソフトの利点は漢字 ROM ボードを必要としないという点であり、従ってノート型でも利用ができる。しかし、利点は欠点でもあり、その結果、これまでの文華のような「フロント文華」として FEP を切り離すことはできなくなった。その代わり「KOA-TechnoMate」にはデータベースが一応付属しているが、これは一種の「差し込み印刷」を利用したものであり、本格的なデータベースとしては実用に耐えないものである。また、プロテクトが施されており、使い勝手は必ずしも良いとは言えないものである。ハードディスクが必須であるのに、ハードディスクに1度インストールすると元のフロッピーは使えなくなり、別のハードにインストールする場合には情報をフロッピーに戻すといった面倒な作業が必要であり、その上インストールの回数も制限されるという極めて時代錯誤的なものとなっている。フロッピーのバックアップすらできないのであるから、困ったものである。

文字種や入力方法等は、「文華13000」と大きな変化はない。ただ、声調付ピンイン入力で大文字入力(Women が wömen としかならない)ができないというバグ(?)がある。なお、「ハングル」などとの混在も可能である。

しかし、私たちにとっては FEP が切り離せないということは、やはり「改悪」としか考えられない。

2.3. 「NiHao-FP Ver3b」(クリエイイト大阪)

これも基本的な機能は「文華13000」とさほど変わらないが、いかんせん「遅い」ということである。これは漢字 ROM ボード方式と、そうでないかの決定的な違いである。これをいくらでも解消する方法としては、文字フォントや辞書を RAM ディスクに置くことであるが、この RAM がプロテクトモードのそれは使えないという不便さがある。(ただし、最新のバージョンでは、改良されているかも知れない)

入力方法は、ピンインのほかに「声形入力」「四角コード入力」などがある。「声形入力」というのは、ピンインの頭文字+漢字の偏旁のそれぞれの頭文字の3つのコードで入力する方法である。たとえば、「張」は zhang+gong+chang のそれぞれの最初の文字を取って zgc となる。ピンインに比べキー入力が少なくすむが、かなり面倒な気もする。「四角コード」とは「四角號碼」のことであり、これは慣れた人には便利なものである。

印刷の面では、ゴシック体や、宋書体、楷書体というふうに書体が豊富なことがあげられる。

「NiHao」も FEP は切り離して、データベースソフト上などでも使えるが、使い勝手は、使用経験がほとんどなくよくわからない。

2.4. 「AllScript」(マキエントプライズ)

これは Mac 上で動く日本語ワープロソフトであるが、マルチリンガル仕様となっており、他の言語のシステムを組み込んで複数の言語の混在が可能である。現在、漢字 Talk6.x と Hanzi, Taiwanese とのハンドリングがサポートされている。ハンドリングされたシステム上では、HyperCard などでも中国語表示ができる。入力方法はピンインのほかに注音字母や四角號碼、字形入力などがある。初めのころは単漢字変換で使いにくかったが、PDS で「CIHUI 入力」が作られ単語変換ができるようになってきている。声調付ピンイン表示は、できないようである。

AllScript 自体は「キーディスク方式」という一種のプロテクトをかけており少々不便であるが、でき上がったシステム上で、他のマルチリンガルサポートソフト(たとえば Nisus など)を使って中国語を入力するという方法もある。

なお、現在の Mac の最新システムは System7.1 である(これに関しては、後に詳しく述べる)が、これに対応した中国語システムはまだ出ていない。

2.5. その他

2.5.1 IBM 対応中国語ソフト

中国本土のコンピュータの主力は IBM 系統であり、当然、中文 DOS で動く中国語ソフトが数多くある。最近いくつかは日本でも入手できるようになっているが、日本語との混在はできないようである。台湾も同じような状況であり、「倚天」「中文之星」といったソフトが知られている。

中国の人はどういうわけか、入力方法に非常にこだわるようである。とにかく「速さ」を重要視して、2つあるいは3つのキー入力可能な「倉頡」という方法や、ローマ字入力でも「双拼」を好むようである。従ってキーボードなども特殊なものになったりするが、汎用性という点や、いくら「速い」といっても将来的には OCR(光学式自動読み取り機)などの普及を考えれば、このことはさほど重要ではないと私は考えている。

なお、つい先ほど、DOS/V 対応の中国語ドライバーが「KOA-TechnoMate」の高電社から発表されたが、これは先の「フロント文華」の DOS/V 版とも呼ぶべきもので、様々なアプリケーションの中での中国語、日本語、英語の混在が可能になるはずである。

2.5.2 「電脳辞林」(エイシス)

これまでのものは中国語ソフトとして開発されたものであるが、この「電脳辞林」は少し発想が変わっており、既存の日本語ワープロソフトの中で中国語を使おうとして考えられたものである。つまり、日本の JIS 漢字を中国の漢字として利用し、足りない部分を「外字」で補おうとするものである。入力方法もピンイン、四角号码、倉頡、カナ代用の注音符号などが考えられているが、いずれも「単語登録」の方法によっている。つまり、「北京」の読みを「ぺきん」ではなく「Peijing31」で登録するわけである。

日本語ワープロの中で外字登録が最も多くできるのはジャストシステムの「一太郎 V4」で、1880字が可能である。これだけあれば、JIS 漢字と併せてほぼ中国漢字は表示できるようである。もちろん、基本は「単語登録」ということであるから、あらゆる日本語ワープロソフトでの応用がきく。「電脳辞林」は市販されたものであるが、同じ原理に基づいた ATOK、VJE 用のピンイン登録辞書が PC-VAN や NIFTY といったパソコン通信ネットに PDS として登録されている。

私は「電脳辞林」も使っているが、FEP は ATOK7 ではなく、エーアイソフトの WX2 を使用している。なぜなら、WX2 の辞書の拡張が最大 4M までという大きなものであるからである。

ただこの「電脳辞林」方式の難点は「一太郎」などのように外字が 1880 字も使えるソフトの中では非常に便利なものであるが、DOS 上では外字は 188 字という制限があり、いささか中国漢字の表示に窮屈であるということである。

しかし、特に日本語ワープロソフトを使いながら中国の古典を引用したりする場合には極めて重宝なものである。

3. Mac の中国語環境

3.1. SYSTEM7.1での可能性

1992年10月13日、アップルコンピュータは「漢字 Talk7リリース7.1」を発表した。これは、平たく言えば「System7.1の日本語版」というものであり、日本語環境の飛躍的向上がなされたことは言うまでもないが、私たちにとって最も関心があるのは、この System7.1なる新しいシステムによって、マルチリンガル（複数言語、多言語）をサポートする基本的な環境が整ったという点である。このマルチリンガルのサポートの最初に登場したのが「漢字 Talk7」ということになるわけである。

3.2. これまでの System

これまでのシステムでは、各言語のシステムは、各言語ごとに別々にローカライズされ、各言語の表示には、「漢字 Talk」「中文 Talk」（Chinese Talk, Hanzhi Talk）「Hangle Talk」とかいった独自のシステム拡張が必要であった。つまり、システムそのものが変更されるわけであり、同じ System6.0.7でも、日本の「漢字 Talk6.0.7」とアメリカの System6.0.7とは内容が異なるということがあったのである。

従って、多言語のサポートとなるとかなり厄介であり、たとえば ScriptMover (MAKI Enterprise Inc.) や ResEdit などを利用して、元のシステムに各言語のシステムをハンドリングするという作業が必要であった。しかもこの場合にもシステムのバージョンが重要であり、異なったバージョンでは使用できなかつたり、システムの一部のリソースレベルの改造を伴うことによる予期できぬリスク（不安定さなど）はユーザ自身が負わなければならなかった。その上、このように作成されたマルチリンガルシステムであっても、使用できるアプリケーションには条件（=ScriptManager に対応しているもの）があり、きわめて限られた

ものであった。特に、中国語という点では先に述べた「AllScript」(マキエンタープライズ)ぐらいであった。

3.3. System7.1について

今回の System7.1 では、WorldScript と、InputMethod という拡張機能を載せることによって、すべて同一の System ファイルで多言語に対応できるようになった。つまり、英語も日本語も中国語も System の変更を加えることなく同じ System で扱えることができるわけである。これは、特に 2 バイト系を扱うものにとっては朗報である。

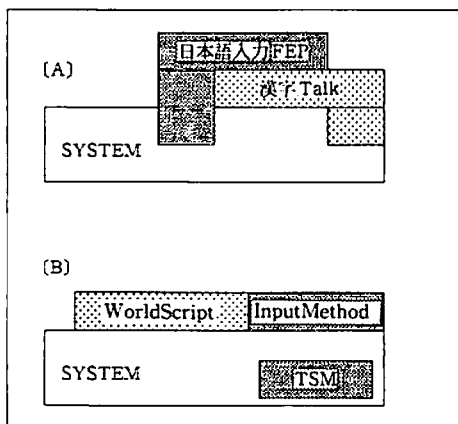
3.3.1. WorldScript

Worldscript とは、これまでの各言語システムのうちの前フォントと InputMethod (入力方法、いわゆる FEP と考えてもいい) を除いた部分、すなわちスクリプトを扱う部分で、たとえば、2 バイト系言語の場合には文字コードの中の 1 バイト目と 2 バイト目かの判断をしたり、1 バイト系ならばアラビア語の右から書く文字の処理などをするものであり、2 バイト系言語の処理用には WorldScript II が、1 バイト言語用には WorldScript I が用意されており、両方使用する場合には同時にインストールすればいいわけである。

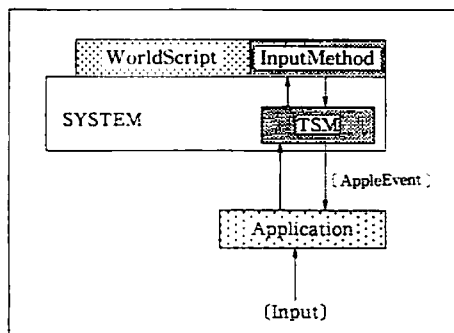
3.3.2. TSM と InputMethod

InputMethod とは、これまで FEP と呼ばれるものに該当するもので、かな漢字変換などの入力処理を行うものである。それをサポートするのが System 内の新しいマネージャである TSM (TextServiceManager) というものであり、InputMethod とアプリケーションのインターフェースをとったり、入力ウィンドーの作成、インライン変換などの処理(将来的には、スペルチェックや検索・置換といった一般的なテキスト処理のサポートも考えられているという)を行うものである。この TSM と InputMethod による入力方式は、これまでの FEP とは、根本的に異なるものである。今までの FEP は、アプリケーションが文字入力を受け取る前に FEP が受け取って変換し、その結果をアプリケーションに渡すという方式であった(ゆえに Front End Processor という)。これに対し新しい入力方法は、これ以外に、アプリケーションが受け取った文字入力を InputMethod に渡して変換し、変換途中の文字列を AppleEvent によってアプリケーションに渡すという方式もサポートしている。つまり、変換は必ずしも Front End で行われるものではないのであり、これが FEP とは呼ばず InputMethod と呼ぶ所以である。「漢字 Talk7」に付属する「ことえり(言選り)」は Apple 純正の日本語入力 InputMethod であり、中国語のそれも近くリリースされてくると思われる。

次ページにこれまでのシステムと System7.1 のシステムを図式して示しておく。



(A)はこれまでのシステム
(B)は System7.1



(新しくサポートされる変換方式)
〔『MacJapan』 92.11より〕

以上のように、System7.1では、共通の System の上に、WorldScript と TSM に対応した複数の各言語の InputMethod、およびフォントを用意すれば、同時に複数の言語を扱うことが可能になったのである。まさに、マルチリンガルシステムの誕生である。このことの持つ意味は大きいものがある。これまで、もともと2バイト系言語は、変換処理やその文字の多さなどからコンピュータで扱うには数々の不便さがあった。さらにその混在となると、一層難しく、限られたアプリケーションの中だけという状況であったが、ここにきてようやくそれが解消され、ワープロのみならずデータ処理、CAI 教育などの本格的なコンピュータの利用に向けた環境が基本的には整ったといえることができるであろう。

4. さまざまな今後の課題

中国語（あるいは広く中国学）のコンピュータ利用は、今ようやくその緒についたばかりと言えるが、さまざまな今後の課題について述べておくこととする。

4.1. 入力方法

中国語の入力方法には、ピンイン入力、部首入力、倉頡入力、注音符号入力、四角号碼入力など多くの方式があり、中国国内だけでも500種類以上も考案されているそうであるが(菅野苓『中国語入力方法の話』1991による)、総じて言えば中国国内では(というか、日本以外では)字形を利用した方法が多いようである。これにはいろんな背景が考えられるが、一つには、できるだけキー入力を少なくしようという意識、つまりできるだけ「多く」「速く」という願望と、今一つにはピンインの普及率ということがあると思われる。しかし、日本における中国語教育・学習は基本的にはピンインを基礎にしており、その教育・学習の効果の面から、またキーボードの国際的規格という面からもアルファベットによる入力、つまりピンイン方式をメインにするのが妥当であると私は考えている。そもそも、いくら「速い」といっても人間の入力には限りがあるし、将来的にはOCRなどの利用を念頭に置くべきであり、「入力量」「速度」というのはそれほど重要なことではないと考えるものである。問題とすべきは、入力方法よりも、その後の処理を如何にすべきであるかということである。次のような意見に耳を傾けるべきである。

中国語文処理の問題点を考えるとき、最大の課題は、このテーマに関わる研究者が、その研究開発のターゲットをどこに見据えていくかにある。単に入力のみを問題とせず、

入力後の大きな目標を見据えた時、中国語の構造を人文科学と計算科学の双方の立場からしっかりと捉える、地味な基礎研究の重要性が見えてくるであろう。その時、現在では、不可能と思われるような課題も、次々に乗り越えられて行くであろう。その最も手短なテーマが、ピンイン入力文章解析型の入力方式であろう。(田嶋一夫「中国語文処理と文化」1990)

なお、このことは日本語の「カナ入力」と「ローマ字入力」との関係でも同様であるように思われる。

4.2. 中国書データベース化の構想とそれに関わる諸問題

これまで、図書の検索にあたって、和書や洋書については国立国会図書館による JAPAN MARC (Machine Readable Cataloging) や、アメリカの議会図書館を中心とする LC (Library of Congress) MARC があり、研究・教育に大きな貢献がなされてきている。一方、中国書(あるいは「漢籍」)については、アメリカやカナダでは CJK (Chinese, Japanese, Korean) システムが開発され、台湾では CHINA MARC が、韓国では KOR MARC が開発されて、それぞれ中国書のデータベース化が始まっているが(この点については山田敬三1992に詳しい)、日本では多くの機関が独自の方法で着手してきているとはいうものの、ハード、ソフト両面において、その互換性、汎用性という点では大きな問題がある。そもそも、日本においては中国書入力のための基本的なフォーマットすら確立していないというのが現状である。

このような状況に対して、1989年秋、日本中国学会第41回全国大会において「漢籍データベース化小委員会」が設置されて検討が開始され、その結果が「中国書データベース化についての答申」(1990.3.11, 『日本中国学会報』第42集所収)としてまとめられた。その後、改めて「中国書データベース化委員会」が設置され、答申の具体化に向けての作業が進められ、目録規則案、フォーマット案、入力規則案からなる「中国書データベース化に関する諸規則(案)」(1991.7.1, 『日本中国学会報』第43集所収)が提案された。これらの具体的な内容については今詳しく述べることはしないが、最終的には国際的なデータ互換を念頭においた JAPAN MARC の中国書版を策定するものとなっており、この場合にも問題となるのは、入力方式、使用漢字、文字コード等である。

ここでは、入力方式については省き、使用漢字、文字コードの問題について以下若干触れておくこととする。

4.2.1 使用漢字(文字数、文字種、文字コード)の問題

まず最初に、この問題は、立場、考え方の違いによって大きく分かれるものであることを明らかにしておく必要がある。

漢字には、いわゆる「正字」以外に「異体字」「俗字」と言われるものがあるが、これらをすべて厳密に区別して処理するか否かという点では見解の相違が見られる。結局、その目的をどこにおくかによるのであるが、検索を主眼とするなら、厳密な区別を要求する必要性はないというのが私の個人的な意見である。極論すれば、そのコンピュータにない漢字は、別の漢字や記号で代用することも考えてよいと思われる。

もちろん、すべての漢字が処理できればそれに越したことはないのであるが、この場合、ではいくらの漢字が必要であろうか。

「中国書データベース化についての答申」によれば、22000字程度を考えている。ちなみに、台湾の情報交換用漢字コード CCCII (Chinese Character Code for Information Interchange) では21004字がコード化され、中国では GB コードの基本集6763字を含めた

21039字が準備されているという。また、1992年に ISO(国際標準化機構)により、DIS-10646 という国際的に統一された文字セットが規格化され、これには、日本の JIS 第1、第2、および1990年に制定された補助漢字もすべて含まれているし、中国の GB コードの簡体字、繁体字や台湾のもの、韓国のハンジャ、アメリカのユニバーサルコード等がすべて含まれ、コード化された漢字の総数は約21000字ということである。

ところで、このように20000字もの漢字を処理しようとすれば、現在のような2バイトでは無理が生ずるわけで、3バイトあるいは4バイトコードの検討も考えられなければならないようになってくる。現に、台湾の CCCII は3バイトコードであるし、「国際的統一漢字コード私案」(丹羽正之1990)なども4バイトを提唱している。確かに国際的に統一されたコードは必要であるが、その場合、これまでのコードとの互換や、機種に依存しないものといったことなども併せて考えて欲しいと願うものである。

さて、こうなってくると、もはや大型コンピュータの世界での話であり、個人ベースのパソコンでは限界が見えてくるわけであるが、パソコンレベルでの開発も同時に進行させていくべきではないかと考えている。

4.3. 中国語教育とコンピュータ

最後に、中国語教育におけるコンピュータ利用ということに少し触れておきたい。

中国語教育のコンピュータ利用(CAI)ということは、現在ほとんど行われていないのが実状であろう。私が知っている範囲では、神戸商科大学(松村文芳氏)ぐらいのものであるが、義務教育でもコンピュータ教育が行われるようになった時代を考えれば、今後、真剣に検討、議論がなされるべきだと考える。もちろん、ハード、ソフト両面の環境の整備が必要であるが、幸いハード面では最近の進歩はめざましいものがあるし、各大学機関においてのハードの導入もかなりのピッチで進んでいるという喜ばしい状況が見られる。問題はソフトであるが、アメリカなどではいくらか開発されているようであるし(たとえば、Mac用中国語発音ソフトとして「Pronunciation Tutor・Speaking-CHINESE」(HyperGlot)があるが、音声も提示され値段の割にはよくできている)、英語を始めとする他の欧米語における先進的な成果を取り入れながら開発を推し進めていく必要があると思われる。もちろん、その前に中国語教育における基本語彙、基本文型、基本文法、グレード等の確立が必要であることは言うまでもないことである。

以上、「中国語とパソコン」について述べてきた。解決されるべき課題は山積みであるが、忘れてはならないことは、あくまでも「人間がパソコンを効果的に利用する」ということであり、その「目的」をはっきりさせることであると思う。目的によって当然、ハード、ソフトが決定されるのであり、その逆ではないということである。

なお、私はパソコンは素人であり、間違った理解をしている部分も多いと思われる。専門の方々の御指教をお願いする次第である。

参考文献

- 藪内清1974『中国の数学』岩波書店
- 本田濟1978『易(上)』朝日新聞社(『中国古典選』1)
- プロビット1989『文華13000Ver.4ユーザーズマニュアル』
- 長瀬真理、西村弘之1986『コンピュータによる文章解析入門—OCP への招待』 オーム社
- 長瀬真理1988『Micro-OCP 文章解析プログラム』 沖田電子技研

- 大西菊太郎1989『BASICによる情報処理入門』 ナカニシヤ出版
 梁敏児1992「中文電腦に関する苦闘」(『The BASIC』No.108技術評論社)
 著者不明1992「これが漢字Talk7だっ!」(『MacJapan』No.42技術評論社)
 著者不明1992「これが漢字Talk7だっ!」(『MacJapan』No.43技術評論社)
 菅野芴1991『中国語入力方法の話』朝日出版社
 丹羽正之1990「漢字コード体系を考える国際統一漢字コード私案」(『文華倶楽部』第4号 プロビット)
 田嶋一夫1990「中国語文処理と文化」(『しにか』Vol.1/No.2大修館書店)
 山田敬三1992「アメリカ・カナダの中国書データベースー日本における中国書データベースとの関連から」(『しにか』Vol.3/No.2大修館書店)
 日本中国学会・漢籍データベース化小委員会1990
 「中国書データベース化についての答申」(『日本中国学会報』第42集)
 日本中国学会・中国書データベース化委員会1991
 「中国書データベース化に関する諸規則(案)」(『日本中国学会報』第43集)
 日本中国学会・中国書データベース化委員会1992
 「中国書データベースの現状と課題」(『日本中国学会報』第44集)

<資料(A)>

A.1.

全知者である覚った人に礼したてまつる。

求道者にして聖なる観音は、深遠な智慧の完成を実践していたときに、存在するものには五つの構成要素があると見きわめた。しかも、かれは、これらの構成要素が、その本性からいうと、実体のないものであると見抜いたのであった。

シャーリプトラよ、

この世においては、物質的現象には実体がないのであり、実体がないからこそ、物質的現象で(あり得るので)ある。

実体がないといっても、それは物質的現象を離れてはいない。また、物質的現象は、実体がないことを離れて物質的現象であるのではない。

(このようにして、)およそ物質的現象というものは、すべて、実体がないことである。およそ実体がないということは、物質的現象なのである。

これと同じように、感覚も、表象も、意志も、知識も、すべて実体がないのである。

シャーリプトラよ、

この世においては、すべての存在するものには実体がないという特性がある。

生じたということもなく、滅したということもなく、汚れたものでもなく、汚れを離れたものでもなく、滅るとということもなく、増すということもない。

それゆえに、シャーリプトラよ、

実体がないという立場においては、物質的現象もなく、感覚もなく、表象もなく、意志もなく、知識もない。眼もなく、耳もなく、鼻もなく、舌もなく、身体もなく、心もなく、かたちもなく、声もなく、香りもなく、味もなく、触れられる対象もなく、心の対象もない。眼の領域から意識の領域にいたるまでことごとくないのである。

(さとりもなければ、)迷いもなく、(さとりがなくなることもなければ、)迷いがなくなることもない。こうして、ついに、老いも死もなく、老いと死がなくなることもないというにいたるのである。苦しみも、苦しみの原因も、苦しみを制することも、苦しみを制する道もない。知ることもなく、得るところもない。それ故に、得るといことがないから、諸の求道者の智慧の完成に安んじて、人は、心を

覆われることなく住している。心を覆うものがないから、恐れがなく、顛倒した心を遠く離れて、永遠の平安に入っているのである。

過去・現在・未来の三世にいます目ざめた人々は、すべて、智慧の完成に安んじて、この上ない正しい目ざめを覚り得られた。

それゆえに人は知るべきである。智慧の完成の大なる真言、大なるさとの真言、無上の真言、無比の真言は、すべての苦しみを鎮めるものであり、偽りがなく真実であると。その真言は、智慧の完成において次のように説かれた。

ガテー ガテー パーラガテー パーラサンガテー ボーディ スヴァーハー

(往ける者よ、往ける者よ、彼岸に往ける者よ、彼岸に全く往ける者よ、さとりよ、幸あれ。)

ここに、智慧の完成の心を終る。

—— 「般若心経」(岩波文庫)より 現代語訳 中村元 ——

A.2

====COMMAND=====

```
*input  comment between '---'.
*words  alphabet  'ー さとり'.
        punctuation with japanese.
*action do wordlist.
        keys sorted by descending frequency and descending length.
*format layout depth 32.
*go
```

====OUTPUT=====

実体	10	大	2	偽	1
物質的現象	8	知	2	恐	1
完成	6	覆	2	見	1
智慧	6	迷	2	滅	1
心	6	目	2	故	1
真言	5	老	2	五	1
さとり	5	一	1	幸	1
往者	4	ラ	1	香	1
人	4	ス	1	次	1
得	4	ヴァ	1	耳	1
離	4	バー	1	終	1
シャ	4	ラ	1	住	1
一	4	ガ	1	諸	1
リ	4	テ	1	上	1
ブ	4	ー	1	触	1
トラ	4	知	1	正	1
成	3	者	1	生	1
要素	3		1	聖	1
ガ	2		1	声	1
テ	2		1	説	1
道	2		1	舌	1
求	2		1	全	1
意	2		1	増	1
志	2		1	鎖	1
感	2		1	同	1
存	2		1	道	1
在	2		1	入	1
象	2		1	鼻	1
対	2		1	味	1
知	2		1	滅	1
彼	2		1	礼	1
岸	2		1		
象	2		1		
表	2		1		
領	2		1		
域	2		1		
安	2		1		
汚	2		1		
覚	2		1		
眼	2		1		
死	2		1		
世	2		1		
制	2		1		

```
TOTAL WORDS READ      = 178
TOTAL WORDS SELECTED  = 178
TOTAL WORDS PICKED    = 178
TOTAL WORDS SAMPLED   = 178
TOTAL WORDS KEPT      = 178
TOTAL VOCABULARY      = 95
```

A.3

====COMMAND=====

```

*input text to 80.
*words alphabet 'ー さとり'.
      punctuation with japanese.
*action do concordance.
      pick words '智慧 実体 物質* さとり 覚*'.
      headwords 'さとり' = 'さとり 覚*'.
      maximum context span L.
*format {context size 3 and indent complete.}
*go

```

====OUTPUT=====

1 全知者である覚った人に礼したてまつる。 求道者に
24 とごとくないのである。 (さとりもなくれば、) 迷いもなく、 (迷
24 もなければ、) 迷いもなく、 (さとりがなくなることもなければ、) 迷
32 て、この上ない正しい目ざめを覚り得られた。 それゆえに人は知るべ
33 完成の大きいなる真言、大きいなるさとりの真言、無上の真言、無比の真言
38 者よ、彼岸に全く往ける者よ、さとりよ、幸あれ。) ここに、智慧

4 要素が、その本性からいうと、実体のないものでありと見抜いたのであ
8 世においては、物質的現象には実体がないのであり、実体がないからこ
8 現象には実体がないのであり、実体がないからこそも、物質的現象で(あ
10 で(あり得るので)ある。 実体がないといつても、それは物質的現象を
10 いない。また、物質的現象は、実体がないことを離れて物質的現象であ
12 的現象というものは、すべて、実体がないことである。およそ実体がない
13 実体がないことである。およそ実体がないこととは、物質的現象な
14 象も、意志も、知識も、すべて実体がないのである。 シャーリプトラ
16 は、すべての存在するものには実体がないという特性がある。 生じた
20 えに、シャーリプトラよ、実体がないという立場においては、物質

2 者にして聖なる観音は、深遠な智慧の完成を實踐していたときに、存在
28 ことがないから、諸の求道者の智慧の完成に安んじて、人は、心を覆わ
31 ます目ざめた人々は、すべて、智慧の完成に安んじて、この上ない正し
33 れゆえに人は知るべきである。智慧の完成の大きいなる真言、大きいなる
35 ら真実であると。その真言は、智慧の完成において次のように説かれた
40 よ、幸あれ。) ここに、智慧の完成の心を終る。 — — — 『

8 ラよ、この世においては、物質的現象には実体がないのであり、実
8 であり、実体がないからこそも、物質的現象で(あり得るので)ある。
10 実体がないといつても、それは物質的現象を離れてはいない。また、物
10 現象を離れてはいない。また、物質的現象は、実体がないことを離れて
11 象は、実体がないことを離れて物質的現象であるのではない。(この
12 (このようにして、)およそ物質的現象というものは、すべて、実体
13 よそ実体がないこととは、物質的現象なのである。 これと同じよ
20 がないという立場においては、物質的現象もなく、感覚もなく、表象も

```

TOTAL WORDS PICKED   =   30
TOTAL WORDS SAMPLED  =   30
TOTAL WORDS KEPT     =   30
TOTAL VOCABULARY     =    4

```

A.4

```

====COMMAND=====
*input text to 80.
*words alphabet 'から ある'.
      punctuation with japanese.
*action do concordance.
      include only phrase '智慧 完成'
      and collocates '*から*' upto 99 '*ある*'.
      maximum context upto '。' and span L.
*format context size 3 and indent complete and print newline '///'.
*go

```

====OUTPUT=====

```

      から ある      2
8   シャーリブトラよ、/// /// この世においては、物質的現象には実
    体がないのであり、実体がないからこそ、物質的///現象で（あり得るの
    で）ある
29
      心///を覆うものがないから、恐れがなく、顛倒した心を遠く離れ
    て、永遠の平安に入っている///のである
      から真実 ある      1
34  いなるさとの真言、///無上の真言、無比の真言は、すべての苦しみを
    鎮めるものであり、偽りがないから真実で///ある
      智慧 完成      6
2   者にして聖なる観音は、深遠な智慧の完成を実践していたときに、存在す
    るものに///は五つの構成要素があると見
28
      故に、得るということ
    がない///から、諸の求道者の智慧の完成に安んじて、人は、心を覆われ
    ることなく住
31
      過去・現在・未来の三世にい
    ます目ざめた人々は、すべて、智慧の完成に安んじて、こ///の上ない正
    しい目ざめを覚り得
33
      智慧の完成の大いなる真言、大いなるさと
    りの真言、///無上の真言、無比の真言は、すべての苦しみを鎮めるも
35
      真言は、智慧の完成において次のように説
40
      智慧の完成の心を終
      本性から ある      1
3   構成要素が、その本性///からいうと、実体のないもので
    あると見抜
      領域から意識 ある      1
23
      眼の領域から意識の領域にいたるまでことごと
    くないのである

```

[資料(B)]

B. 1.

<T 华语拼字妙法>

<K 41>

<P 84>

<L 1>

天国。天上。地下。地方。田地。比方。方能。比比 看。住处。各处。到处。好处。
天国 近 了，你们 要 悔改。上帝 在 天上，也 能 看见 地下人 所做的事。中
国 地方 比 外=
国 地方 大。中国人 比 外国人 多。他 来 了，我们 方 能 做 这件事情。请
你 比比 看，这
两件 衣裳，是 哪一件 好。你 现在 没有 住处，请 到 我家里 来 住。各处
有 各处 所出的=
东西，各处 有 各处的话。现在 到处 有 讲天国道理的人。讲道理的人 应该 多
用 比方，
叫 人 容易 明白。我 来 了，恐怕 你 没有 甚么 好处，我家的 田地 不多，
但 比 你家 多 一些。

<K 42>

<P 86>

<L 1>

早已。早上。早起。迟早。早饭。中饭。天 不 早 了。早 晓得。来 得 迟。起工。
起头。
起见。觉得。睡觉。这本书 我 早已 读 过了。早 晓得 这样子，我 大 不该 起
早 来。你
不要 怕 他，迟早 些 必定 是 要 来 的。我 没有 吃 早饭，所以 要 早些
吃 中饭。天 不 早 了，
你 快些 起来 罢。你 起工 的 时候 太 早 了。为甚么 起见 不用 他，就是
因为 他 来 得 迟。
起头 不过 五个人 来，后来 的 人 就 多 了。你 觉得 自己的罪，耶苏 就 肯
饶恕 你 了。你
睡觉 去 罢，明天 早些 起来。早起迟睡。明天的事情 甚 多，请 你 不要 迟到
。他 现在
睡觉，你 不要 叫 他 起来。他 明天 早上 要 来。

<K 43>

<P 88>

B. 2.

< 中国語用語索引[CONCORDANCE]作成のためのコマンド例 >

```
*INPUT TEXT HYPHEN '='.  
REFERENCES COCOA '< to >'.  
*WORDS ALPHABET ':33 :35 :36 :37 :38 :39 :40 :41 :43 :45 :47 :48 :49 :50  
:51 :52 :53 :54 :55 :56 :57 :58 :60 :61 :62 :64 :91 :92 :93 :94  
:95 :96 :123 :124 :125 :126 :160 :161 :162 :163 :164 :165 :166  
:167 :168 :169 :170 :171 :172 :173 :174 :175' WITH JAPANESE.  
PUNCTUATION '。 , ? '.  
*ACTION DO CONCORDANCE.  
PICK WORDS 'ト/ ニク1 ネソクW キ7コJ'.  
KEYS SORTED BY ASCENDING FREQUENCY.  
REFERENCES PRINT T=4 LEFT WITH '-'.K=3 LEFT WITH '-'.P=3 RIGHT  
WITH '-'.L=2 RIGHT.  
MAXIMUM CONTEXT LEFT UPTO '。 , ?' AND RIGHT UPTO '。 , ?'.  
*FORMAT CONTEXT SIZE 2 AND INDENT COMPLETE.  
LAYOUT DEPTH 40.  
*GO
```

< 中国語語彙索引[INDEX]作成のためのコマンド例 >

```
*INPUT TEXT HYPHEN '='.  
REFERENCES COCOA '< to >'.  
*WORDS ALPHABET ':33 :35 :36 :37 :38 :39 :40 :41 :43 :45 :47 :48 :49 :50  
:51 :52 :53 :54 :55 :56 :57 :58 :60 :61 :62 :64 :91 :92 :93 :94  
:95 :96 :123 :124 :125 :126 :160 :161 :162 :163 :164 :165 :166  
:167 :168 :169 :170 :171 :172 :173 :174 :175' WITH JAPANESE.  
PUNCTUATION '。 , ? '.  
*ACTION DO INDEX.  
KEYS SORTED BY ASCENDING FREQUENCY.  
REFERENCES PRINT K=3 LEFT WITH '-'.P=3 RIGHT WITH '-'.L=2 RIGHT.  
*FORMAT LAYOUT NO PAGES AND LINES 0 BELOW ENTRIES.  
CONTEXT INDENT COMPLETE.  
*GO
```

B. 3.

	田地	2
华语-41 - 84- 1	田地	
华语-41 84- 6	我家的 田地 不多	
	让	3
华语-53 -108- 6	让 他们两人 去 分 罢	
华语-55 -112- 4	不要 让 他们 跟 你 去	
华语-60 -122- 2	若是 让 你的儿子 这样 做	
	成功	4
华语-43 - 88- 1	成功	
华语-43 - 88- 2	我 要 做 成功 这件事情	
华语-43 - 88- 3	成功	
华语-53 -108- 6	他 自 分 可 以 成 功 么	
	晓得	4
华语-42 - 86- 1	早 晓得	
华语-42 - 86- 2	早 晓得 这样子	
华语-34 - 70- 4	他 晓得 必定 做不好	
华语-34 - 70- 5	你 晓得 上帝 么	

TOTAL WORDS READ	=	3136
TOTAL WORDS SELECTED	=	3136
TOTAL WORDS PICKED	=	13
TOTAL WORDS SAMPLED	=	13
TOTAL WORDS KEPT	=	13
TOTAL VOCABULARY	=	4

B. 4-1

劝勉 4 99-200-6, 46-94-1, 46-94-2, 50-102-3
 然后 1 70-142-5
 让 3 53-108-6, 55-112-4, 60-122-2
 让位 2 55-112-2, 55-112-5
 饶恕 8 40-82-5, 42-86-5, 44-90-7, 38-78-1, 38-78-2, 38-78-2, 38-78-4,
 38-78-4
 热 7 47-96-5, 33-68-2, 33-68-6, 36-74-6, 57-116-2, 87-176-3, 87-176-4
 热的时候 1 36-74-5
 热了 1 84-170-5
 热水 2 63-128-4, 63-128-4
 热心 1 33-68-2
 人 53 99-200-2, 99-200-2, 99-200-6, 10-22-6, 10-22-6, 5-12-3, 5-12-4,
 15-32-3, 17-36-4, 18-38-5, 18-38-6, 18-38-6, 20-42-5, 21-44-3, 23-48-6,
 24-50-5, 45-92-2, 46-94-1, 46-94-5, 47-96-6, 41-84-6, 43-88-4, 44-90-5,
 34-70-6, 36-74-5, 37-76-5, 38-78-2, 50-102-7, 52-106-5, 55-112-3,
 58-118-4, 60-122-5, 91-184-7, 91-184-7, 93-188-4, 93-188-4, 94-190-2,
 95-192-6, 61-124-6, 62-126-2, 65-132-4, 66-134-6, 68-138-2, 68-138-2,
 68-138-7, 69-140-1, 73-148-4, 76-154-4, 77-156-6, 80-162-6, 85-172-4,
 90-182-2, 98-198-6
 人吃人了 1 75-152-6
 人的 3 48-98-4, 48-98-5, 64-130-5
 人的手 2 12-26-5, 66-134-6
 人的心 1 64-130-5
 人的主 1 37-76-5
 人家 2 100-202-7, 93-188-6
 人口 5 9-20-4, 5-12-2, 5-12-4, 35-72-5, 58-118-6
 人人 5 20-42-6, 46-94-3, 50-102-3, 62-126-1, 79-160-6
 人上人 2 65-132-2, 65-132-6
 人那 2 23-48-2, 23-48-6
 人所爱的 1 53-108-4
 人所恶的 1 53-108-4
 人心 1 98-198-2
 人缘 2 75-152-1, 75-152-4
 人走 1 71-144-5
 认不得 2 17-36-4, 17-36-4
 认错 1 85-172-3
 认得 10 17-36-1, 17-36-2, 17-36-3, 17-36-4, 17-36-4, 34-70-2, 55-112-6,
 55-112-8, 85-172-2, 85-172-3
 认识 1 17-36-1
 认识的字 1 32-66-2
 认字 2 17-36-1, 17-36-2
 软 1 99-200-4
 软弱 2 99-200-1, 99-200-4
 晒不干 1 63-128-6
 晒干 1 63-128-1
 晒干了 2 63-128-3, 63-128-7
 晒晒 1 63-128-1
 晒晒干 2 64-130-2, 74-150-4
 晒台 1 63-128-1
 晒台上 1 63-128-3
 山上 1 72-146-1
 山水 1 72-146-5
 日本国的 1 77-156-3
 日后 1 73-148-6
 日里 3 84-170-1, 84-170-7, 87-176-2
 日夜的 1 87-176-2
 日夜的时候 1 84-170-6
 容易 2 31-64-1, 31-64-6
 容易 11 41-84-6, 31-64-1, 31-64-2, 31-64-3, 39-80-6, 35-72-4, 95-192-3,
 67-136-4, 68-138-2, 72-146-3, 84-170-5
 容易的 1 70-142-2
 容易算 1 80-162-4
 容易做 1 71-144-4

B. 4-2

听从	1	81-164-1
听见	5	33-68-1, 33-68-3, 52-106-5, 54-110-2, 81-164-2
听说	5	44-90-4, 33-68-2, 33-68-3, 51-104-4, 75-152-6
听听	1	75-152-3
听听看	2	33-68-1, 33-68-2
听闲	1	52-106-4
停一停	2	97-196-1, 97-196-4
停的	1	97-196-7
停住	1	97-196-1
停住了	1	97-196-4
通不通	2	74-150-1, 74-150-4
通气	1	77-156-6
通商	1	74-150-6
通行	1	74-150-1
通行的	2	97-196-6, 74-150-4
通用	1	74-150-1
通用的	1	74-150-4
同	25	99-200-3, 25-52-2, 25-52-3, 25-52-5, 25-52-6, 28-58-3, 47-96-6, 36-74-2, 36-74-2, 35-72-3, 35-72-6, 35-72-6, 49-100-2, 49-100-3, 49-100-5, 53-108-3, 58-118-3, 58-118-5, 94-190-6, 95-192-3, 74-150-2, 74-150-5, 75-152-2, 78-158-2, 81-164-5
同不同	2	25-52-1, 25-52-5
同年	1	76-154-1
同年的	1	76-154-5
同心	2	25-52-1, 25-52-5
钢的	1	61-124-6
钢钱	2	61-124-1, 61-124-2
铜头	2	14-30-4, 48-98-5
头一回	1	85-172-4
头顶上	1	73-148-1
头上	1	73-148-2
头一个	1	73-148-3
头一个月	2	14-30-1, 14-30-5
头一回	1	100-202-6
土	2	14-30-1, 14-30-7
土音	1	14-30-6
土中	2	97-196-1, 97-196-6
外国人	1	96-194-3
外国人	4	41-84-2, 72-146-3, 84-170-6, 84-170-7
外国先生	7	41-84-3, 44-90-7, 39-80-5, 52-106-7, 95-192-2, 61-124-2, 63-128-4
外面	1	97-196-5
外人	2	86-174-1, 86-174-3
外头的衣裳	2	15-32-1, 15-32-6
外衣	6	15-32-1, 15-32-2, 15-32-3, 20-42-7, 33-68-4, 64-130-2
为的	1	63-128-7
为己	2	15-32-1, 15-32-6
为记	2	62-126-2, 87-176-4
为人	1	65-132-3
为人好	2	19-40-1, 19-40-2
为甚	1	78-158-3
我	4	19-40-1, 19-40-2, 19-40-2, 20-42-4
脱了	1	75-152-4
脱落下来	18	19-40-1, 19-40-2, 19-40-7, 21-44-2, 29-60-2, 30-62-6, 40-82-2, 47-96-4, 47-96-7, 42-86-4, 33-68-7, 51-104-6, 52-106-5, 54-110-2, 65-132-6, 67-136-5, 76-154-3, 84-170-2
脱完	1	64-130-3
脱完了	1	64-130-3
脱完	2	64-130-1, 64-130-6
脱完	1	64-130-2
脱完	1	64-130-1
脱完	2	50-102-1, 50-102-4
脱完	2	50-102-1, 50-102-2