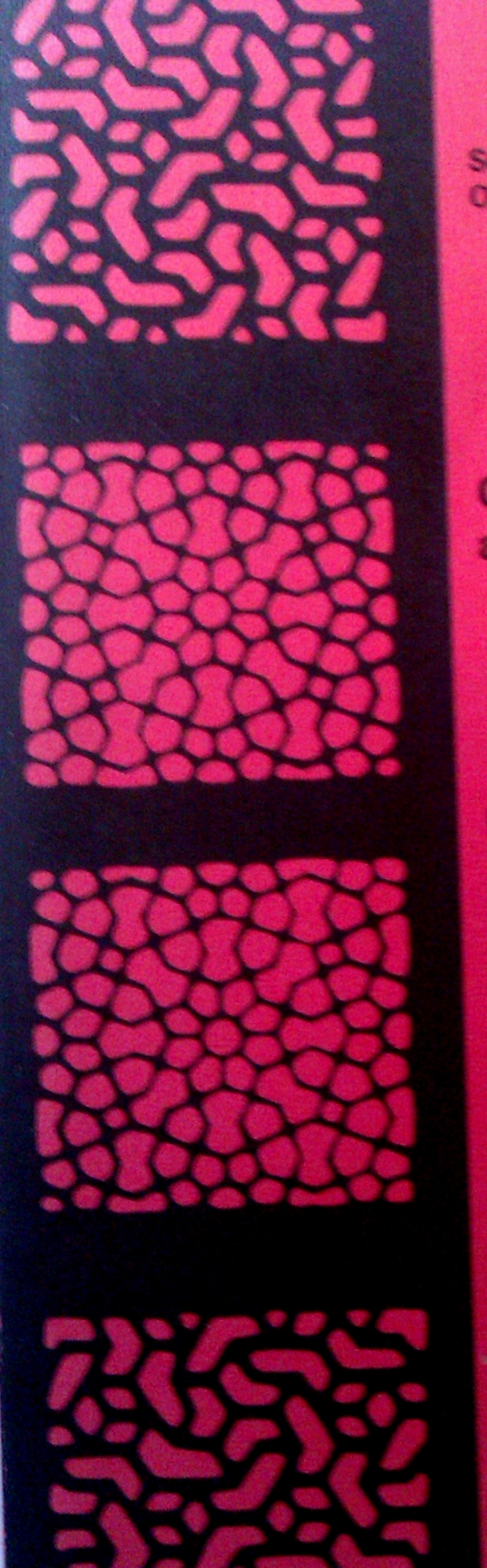

South Asia Series
Occasional Paper No. 40

Calcutta, Bangladesh, and Bengal Studies

1990 Bengal Studies Conference Proceedings

Edited by Clinton B. Seely

ASIAN STUDIES CENTER

MICHIGAN STATE UNIVERSITY

WORDPROCESSING, DATABASES, AND SPREADSHEETS IN BENGALI

Anisur Rahman

INTRODUCTION

Wordprocessing in Bengali is no longer a personal thing among a few overenthusiastic people. These days Bengali wordprocessing is being used not only for personal correspondence but also for various publishing works on an international scale. A wider use of the Bengali language on personal computers has been facilitated by the ability to use "soft fonts" in wordprocessing programs on certain printers.

Today personal computers provide the power once available only from mainframe and minicomputers. With hardware and software working together, the personal computers form the core of scientific, engineering, desktop publishing, and other numerous applications. The major hardware components contributing to this power are the microprocessor chips (Intel 808x and 80y86 family for IBM and compatible personal computers), fixed disk and/or other bulk storage devices, the memory spaces (random access memory, RAM), the pointing devices, scanners, and for some applications a floating point coprocessor.

The advantage of using personal computers is manifold. As a replacement of mainframes they offer increased independence, flexibility, and cost savings. As a replacement of manual efforts they offer tremendous time savings and increased accuracy. The revolutionary idea of desktop publishing on a personal computer was implemented around 1985. With the aid of some special programming techniques, e.g., the graphical device interface and dynamic data exchange ability, the personal computers can do in minutes what it once took hours to accomplish with traditional typesetting, layout, and pasteup methods.

Up until now the use of Bengali and other Asian and Arabian languages on personal computers solely depended on the ability of downloading of respective soft fonts on the printers. This feature has to be utilized until the read only memory (ROM) chips containing the fonts of Bengali (and other language) are available.

In order to make personal computers more useful to a society, it is time to step forward and introduce other applications such as databases and spreadsheet programs in Bengali. This can be done by developing appropriate software which would work on the existing hardware until

Bengali ROM chips are available. To develop one's own software is probably the most desirable but would be a complicated project and would require considerable amount of time, funding, and resources. An alternative way is to use the Bengali fonts in existing commercial software. By doing so, most of the facilities provided by a suitable software can be used in Bengali, and almost all the difficulties of the development stage can be avoided. The task then is how to create a suitable font set and find a way of using it with a suitable commercial software.

In the present article the basics of wordprocessors, databases, and spreadsheet programs will be reviewed, with a few examples using Bengali fonts in a commercially available software.

WORDPROCESSOR

A wordprocessor is a computer program that takes an ASCII (American Standard Code for Information Interchange) text, usually by means of the keyboard input, and then formats the text in the desired way for printing or screen display. There are many different algorithms for wordprocessing programs. In the simplest form a "wordprocessed file" can be envisioned as a big matrix that contains all the text and commands entered by the user from the keyboard. During the processing time, the program translates the wordprocessor commands into the specific printer commands and sends them to the printer via an appropriate protocol known as a printer driver.

The ASCII is a set of 256 characters used by the computer to communicate with the peripherals (e.g., a printer), monitor, disks, as well as with its memory. Out of these 256 characters 0-31 are nonprintable control codes; 32-126 are the text characters which include the special characters, numerals, and upper and lower case letters; 127-255 are known as higher (or extended) ASCII. The higher ASCIIs can be used for control codes (127-160) and for additional characters (161-255). A list of ASCII characters is shown in Table 1.

The main difference between using English and using Bengali on a computer lies in the total number of characters required for these two languages. In English there are fifty-two lower and upper case letters, ten numerals and thirty-three special characters (including the space). A standard keyboard supplies all these characters plus a few special keys. In Bengali there are fifty letters including eleven vowels, and the same number of numerals and special characters as in English. (Bengali and other Asian languages, moreover, have no separate upper case and lower

case letters.) If this were the only difference between Bengali and English, the disparity would be considered relatively minor. The biggest difference, however, arises from the fact that Bengali uses a lot of compound letters or ligatures. These compound letters are somewhat analogous to diacritical marks and make computing in Bengali radically different from computing in English. Compound letters are present not only in Bengali but also in many other languages, for instance, Arabic, Hindi, Sinhalese, and Assamese, which uses the Bengali script. Because of these compound letters in Bengali, about 135 characters (or more, depending on orthographic style) are required in order to spell most words. Table 2 displays the keyboard layout that I have used on a 101-key AT keyboard, and Table 3 shows the higher ASCII assignments. As is seen from Tables 2 and 3, most of the whole letters are kept in the main keyboard, and the compound letters and half letters are kept in the higher ASCII portion of the font. This is so done because the rule of compounding characters is not uniform. From Table 3 it is seen that the letters are assigned at a somewhat random fashion, the reason for which is explained below.

Because of the difference in the total number of characters, it is a bit inconvenient to use the same keyboard for both English and Bengali. One way of getting around this problem is to use the higher ASCII portion (128-255) of the ASCII set. Actually this is being practiced in all of today's Bengali applications on the computer as well as in all other foreign languages that require more characters than are available on a standard keyboard. In order to access the higher ASCII characters one has to use multiple keystrokes on the numeric keypad. That is, if I want to get the character which is assigned to ASCII 163 (see Table 3), I have to hold the "Alt" key and then type 163 on the keypad. Thus four keystrokes are needed to produce a higher ASCII character. This may seem to be a little awkward, but if one compares this with other standard wordprocessors like WordPerfect, one will find that entering a Greek letter, for example, takes as many as seven to nine keystrokes! On the numeric keypad pressing one three-digit sequence is no different from pressing any other three-digit sequence. That is the reason for the somewhat random assignment of the higher ASCII characters. However, the keyboard assignments were done so as to keep the phonetic similarity of a letter as close as possible in both the languages. For example, velar stop (first letter of the Bengali consonants) was assigned to "k," and so on.

FONTS AND FONT FORMATS

A font is a set of characters or letters designed for a specific model of printer or display device. The word font refers to the shape and size

of that set. In normal applications it is the printer that is considered first. A pair of matching fonts is created: one for the printer and one for the screen. The screen font depends on the screen resolution, that is, on the type of graphic adapter being used. The types of printer fonts one can use vary with the kind of printer involved. For most printers there are three types of printer fonts available:

1. ROM fonts: ROM (read only memory) fonts are also called built-in fonts. These are the fonts installed in the ROM of the printer by the manufacturer and, therefore, depend on the specific printer;

2. cartridge fonts: These fonts are plugged into the printer. They come in several different styles and point sizes;

3. soft fonts: The soft fonts or downloadable fonts are those one can transfer from the computer to the printer. Generally, the ability to download fonts gives one the greatest variety of fonts. However, since these fonts are downloaded into the RAM of the printer, an appreciable amount of RAM is necessary in order to hold the font(s) to be downloaded.

Another factor that determines the ability of downloading of the soft fonts is the format of the fonts used by a particular printer. For instance, there are slight differences between the formats used by HP Laserjet+ and HP Series II printers. For most symbol sets, i.e., fonts, the Series II allows ASCII 0-31 and 128-159 to be downloaded whereas the Laserjet+ does not. The sets of Bengali fonts currently developed, therefore, do not define characters in those ranges.

DATABASES AND SPREADSHEETS

Once the Bengali character sets are available, a suitable word-processing program can use them. Other useful programs, which can take advantage of the Bengali characters on a computer, are databases, spreadsheets, graphics packages, and other programs which are in widespread use. Such specialized use of Bengali on computers is also very timely and well justified.

Databases

A database is a collection of data organized in a logical, consistent order that allows simple and flexible retrieval and updating. Each logical record is divided into fields, and each field contains particular information.

A good database program can be used to do the following tasks:

1. searching or querying the database to obtain information about the contents;

2. extracting subsets of the data based on some specific criteria;

3. performing statistical manipulations on the data for analysis and decision making;

4. sorting and merging alphabetically, numerically, in ascending or descending order, by names or locations, or by other categories specified by the user;

5. creating and printing out reports using the data tailored to specific purposes.

For instance, one might set up a database as a list of students, their addresses, year in the program, status of the comprehensive exam, etc. Then the database could be used to extract information and make a report about a group of students based on given criteria. A sample database of names and addresses is listed under Table 5.

Spreadsheet

A spreadsheet—also called worksheet—is the primary document one uses for storing and manipulating data. A good spreadsheet program provides a number of spreadsheet tools needed for projections, calculations, statistical and mathematical analysis, and decision making and reporting. In addition, it also provides means of creating graphical presentation of data. Today's databases and spreadsheets are used in everyday life almost everywhere in the business world and other places.

CONCLUSIONS

The Bengali fonts that appear in this paper were used with applications that run under Microsoft Windows. The advantage of being able to use a font under Windows is that all Windows-based applications can share it. For instance, Tables 2, 3, and 4 were created with PageMaker; the sample database (see Table 5) was created with Microsoft Excel. This ability to share fonts among applications provides a means of employing Bengali in almost every aspect of computer use.

Table 1

Ctrl	Dec	Hex	Char	Code
^@	0	00		NUL
^A	1	01	☺	SOH
^B	2	02	☻	STX
^C	3	03	♥	ETX
^D	4	04	♦	EOT
^E	5	05	♣	ENQ
^F	6	06	♠	ACK
^G	7	07	•	BEL
^H	8	08	◘	BS
^I	9	09	○	HT
^J	10	0A	◙	LF
^K	11	0B	♂	VT
^L	12	0C	♀	FF
^M	13	0D	♪	CR
^N	14	0E	♫	SO
^O	15	0F	☼	SI
^P	16	10	►	DLE
^Q	17	11	◄	DC1
^R	18	12	↕	DC2
^S	19	13	‼	DC3
^T	20	14	¶	DC4
^U	21	15	§	NAK
^V	22	16	▬	SYN
^W	23	17	↨	ETB
^X	24	18	↑	CAN
^Y	25	19	↓	EM
^Z	26	1A	→	SUB
^[	27	1B	←	ESC
^\	28	1C	∟	FS
^]	29	1D	↔	GS
^^	30	1E	▲	RS
^_	31	1F	▼	US

Dec	Hex	Char
32	20	
33	21	!
34	22	"
35	23	#
36	24	$
37	25	%
38	26	&
39	27	'
40	28	(
41	29	)
42	2A	*
43	2B	+
44	2C	,
45	2D	-
46	2E	.
47	2F	/
48	30	0
49	31	1
50	32	2
51	33	3
52	34	4
53	35	5
54	36	6
55	37	7
56	38	8
57	39	9
58	3A	:
59	3B	;
60	3C	<
61	3D	=
62	3E	>
63	3F	?

Dec	Hex	Char
64	40	@
65	41	A
66	42	B
67	43	C
68	44	D
69	45	E
70	46	F
71	47	G
72	48	H
73	49	I
74	4A	J
75	4B	K
76	4C	L
77	4D	M
78	4E	N
79	4F	O
80	50	P
81	51	Q
82	52	R
83	53	S
84	54	T
85	55	U
86	56	V
87	57	W
88	58	X
89	59	Y
90	5A	Z
91	5B	[
92	5C	\
93	5D	]
94	5E	^
95	5F	_

Dec	Hex	Char
96	60	`
97	61	a
98	62	b
99	63	c
100	64	d
101	65	e
102	66	f
103	67	g
104	68	h
105	69	i
106	6A	j
107	6B	k
108	6C	l
109	6D	m
110	6E	n
111	6F	o
112	70	p
113	71	q
114	72	r
115	73	s
116	74	t
117	75	u
118	76	v
119	77	w
120	78	x
121	79	y
122	7A	z
123	7B	{
124	7C	|
125	7D	}
126	7E	~
127	7F	⌂†

Dec	Hex	Char
128	80	Ç
129	81	ü
130	82	é
131	83	â
132	84	ä
133	85	à
134	86	å
135	87	ç
136	88	ê
137	89	ë
138	8A	è
139	8B	ï
140	8C	î
141	8D	ì
142	8E	Ä
143	8F	Å
144	90	É
145	91	æ
146	92	Æ
147	93	ô
148	94	ö
149	95	ò
150	96	û
151	97	ù
152	98	ÿ
153	99	Ö
154	9A	Ü
155	9B	¢
156	9C	£
157	9D	¥
158	9E	₧
159	9F	ƒ

Dec	Hex	Char
160	A0	á
161	A1	í
162	A2	ó
163	A3	ú
164	A4	ñ
165	A5	Ñ
166	A6	ª
167	A7	º
168	A8	¿
169	A9	⌐
170	AA	¬
171	AB	½
172	AC	¼
173	AD	¡
174	AE	«
175	AF	»
176	B0	░
177	B1	▒
178	B2	▓
179	B3	│
180	B4	┤
181	B5	╡
182	B6	╢
183	B7	╖
184	B8	╕
185	B9	╣
186	BA	║
187	BB	╗
188	BC	╝
189	BD	╜
190	BE	╛
191	BF	┐

Dec	Hex	Char
192	C0	└
193	C1	┴
194	C2	┬
195	C3	├
196	C4	─
197	C5	┼
198	C6	╞
199	C7	╟
200	C8	╚
201	C9	╔
202	CA	╩
203	CB	╦
204	CC	╠
205	CD	═
206	CE	╬
207	CF	╧
208	D0	╨
209	D1	╤
210	D2	╥
211	D3	╙
212	D4	╘
213	D5	╒
214	D6	╓
215	D7	╫
216	D8	╪
217	D9	┘
218	DA	┌
219	DB	▒
220	DC	▄
221	DD	▌
222	DE	▐
223	DF	▀

Dec	Hex	Char
224	E0	α
225	E1	ß
226	E2	Γ
227	E3	π
228	E4	Σ
229	E5	σ
230	E6	µ
231	E7	τ
232	E8	Φ
233	E9	Θ
234	EA	Ω
235	EB	δ
236	EC	∞
237	ED	φ
238	EE	ε
239	EF	∩
240	F0	≡
241	F1	±
242	F2	≥
243	F3	≤
244	F4	⌠
245	F5	⌡
246	F6	÷
247	F7	≈
248	F8	°
249	F9	∙
250	FA	·
251	FB	√
252	FC	ⁿ
253	FD	²
254	FE	■
255	FF	

† ASCII code 127 has the code DEL. Under DOS, this code has the same effect as ASCII 8 (BS).
The DEL code can be generated by the CTRL + BKSP key combination.

Table 2
Keyboard layout
(101-Key AT Type)

~	!	@	#	$	%	^	&	*	(	)	_	+
্	!	ঙ্	#	৪	%	ৗ	ৎ	*	(	)	ঃ	+
`	1	2	3	4	5	6	7	8	9	0	-	=
‘	১	২	৩	৪	৫	৬	৭	৮	৯	০	-	=
Q	W	E	R	T	Y	U	I	O	P	{	}	|
ঢ	ঙ	এ	ড়	ঠ	ঊ	ূ	ী	য়	থ	{	}	।
q	w	e	r	t	y	u	i	o	p	[	]	\
ত	ত্র	ই	র	ট	ঐ	ু	ি	ও	প	[	]	ং
	A	S	D	F	G	H	J	K	L	:	"	
	া	শ	ড	ঢ়	ঘ	ঝ	ছ	ক্ত	ল	:	"	
	a	s	d	f	g	h	j	k	l	;	’	
	অ	স	দ	ফ	গ	হ	য	ক	ল	;	’	
	Z	X	C	V	B	N	M	<	>	?		
	ষ	ক্স	ক্ষ	ঈ	ধ	ণ	ম	ে	্য	?		
	z	x	c	v	b	n	m	,	.	/		
	জ	চ	থ	ভ	ব	ন	ম	,	.	/		

Table 3
Higher ASCII assignments

Dec	Char	Dec	Char	Dec	Char	Dec	Char	Dec	Char
161	ণ্ড	169	ক্ষ	177	ঘ	185	চ	193	´
162	ণ্ঢ	170	গ	178	ঙ	186	ল	194	`
163	ঞ্জ	171	ঘ	179	ক	187	ট	196	ক্র
164	ঞ্ছ	172	ঙ	180	ঙ	188	ক	197	হ্ন
165	ঞ্ঝ	173	দ	181	ত	189	ত	198	ক্ল
166	থ্র	174	ন	182	ন	190	্র	199	ষ্ণ
167	উ	175	প	183	শ	191	়	200	ট্ট
168	ঊ	176	ভ	184	ব	192	ৈ	201	ণ্ব
				195	ৎ				

Table 4
A few words with compound letters

অগ্নী	উপলব্ধি	কল্পনা	খন্দক	সুগন্ধ	স্কন্ধ
অন্তর	উদ্ভাবন	কলঙ্ক	খণ্ডন	সংক্রান্ত	সঙ্ক্রান্ত
স্খলিত	অযোদ্ধ্যা	উচ্ছাস	কঙ্কে	খদ্দর	সম্পন্ন
সৃষ্টি	অক্টোপাস	উন্মাদ	কন্দর্প	গিন্নী	নিম্নজ
স্কটল্যান্ড	অরুন্ধতী	উদ্ভীন	কান্ত	জব্দ	তুরস্ক
দম্ভোক্তি	অনুচ্ছ	আহ্লাদ	কন্টক	জন্মান্ধ	তত্ত্বজ্ঞ
ধ্বংস	আম্পর্থা	আব্বা	উন্মুখ	সন্ধান	স্বর্ণ
মন্দির	উৎকৃষ্ট	প্রাপ্ত	পলাস	ফাল্গুন	বসুন্ধরা
সুয়ম্বরা	স্বপ্নাদ্য	প্রপঞ্চ	বাঞ্ছা	ঝঞ্ঝা	ইঞ্জিনীয়ার
পল্লব	কম্পাউন্ডার	কোষ্ঠী	বিচ্ছেদ	ট্যাক্স	জ্যোতিশ্চক্র
শ্রদ্ধেয়	দুঃস্পর্শ	দুন্দুভি	দৃষ্ট	পন্ডিত	স্ত্রী
দ্বীপপুঞ্জ	রন্ধ্রীয়	লম্পট	লুন্ঠিত	শিল্পযন্ত্র	পৃষ্ঠা
জ্বলন্ত	স্ক্রু	বোল্ট	হিস্টরী	হৃফিপদ্ম	উষ্ণ
ল্যাট্টু	শ্বষ্ট	পঞ্চাশ	ব্রাহ্মন	জিজ্ঞাসা	ফেরেস্তা

Table 5
A sample database of name and addresses

LAST	FIRST	ADDR	CITY	ZIP
আহমেদ	মি: লস্কর	811 N. 15th St., #10	Milwaukee	WI 53233
আহমেদ	ড: রফিক	Dept. of Geog, UW La Crosse	La Crosse	WI 54601
আলম	মি: বদিউল	2235 Woodview Ct., #24	Madison	WI 53713
আমিন	মি: নুরুল	1521 W Kilbourn Ave., #408	Milwaukee	WI 53233
চৌধুরী	ড: আ. রহিম	7315 W. Marine Dr.	Brown Deer	WI 53223
চৌধুরী	মি: নবিউল্লাহ্	809 F, Eagle Heights	Madison	WI 53705
দৌলা	ড: আশ্রিফ উদ	4541 South 23rd St., #4	Milwaukee	WI 53221
গোমেজ	ড: ন্যান্সী	2915A N. Weil St.	Milwaukee	WI 53212
গোমেজ	মি: জেরী	724 East Hampton	Whitefsh B.	WI 53217
গুবার্গ	ড: মার্টিন	1454 Mary Copa Dr.	Oshkosh	WI 54904
হোসেন	মি: আজাদ	1621 W Wells Ave., #416	Milwaukee	WI 53233
হোসেন	ড: মাহমুন	UW Stout	Menomonie	WI 54751
হোসেন	মি: আইয়ুব	3726 W. Juniper Ct.	Milwaukee	WI 53209
খান	ড: জিল্লুর	Pol. Sc. Dept., UW Oshkosh	Oshkosh	WI 54901
মঈন	মি: আরিফ	2102 University Ave	Madison	WI 53705
মুসা	মি: মোহাম্মদ	2545 N. Maryland Ave.	Milwaukee	WI 53211
নবী	মি: দেওয়ান	540 Vernon Place	Elm Gorve	WI 53122
নেওয়াজ	মি: মুরাদ	3209 W. Wisconsin Ave.	Milwaukee	WI 53208
রাব্বানী	মি: নাসিম	4484 W. Deer Run Dr.	Brown Deer	WI 53223
রহমান	আনিসুর	722 N 13th Street	Milwaukee	WI 53233
রহমান	মি: মোফাখ্খার	3568 N. Oakland Ave.	Milwaukee	WI 53211
রশীদ	মি: মোহাম্মদ	1283 Home Ave.	Menasha	WI 54952
রউফ	ড: এম. এ.	1630 Cliffview Dr.	Oshkosh	WI 54901
সাহা	মি: অঞ্জন	3237 N. Okland St.	Milwaukee	WI 53211
সুলায়মান	মি: মোহাম্মদ	1706 Fairway	Oshkosh	WI 54901
উদ্দীন	ড: জমির	613 Pioneer Tower, 1 Unv. Pla	Platteville	WI 53818