

# Letters with Alternative Basic Shapes

Earl M. Herrick

In many written languages there are letters each of which may be embodied by marks having more than one basic shape. For each occurrence of such a letter, the shape of the mark used to embody it is normally selected according to the circumstances in which the letter occurs. Thus, some alternative basic shapes are appropriate to certain places in a word or another part of a text; some are used by different dialects; some belong to different coscripts (subdivisions of a script that each have basic shapes for all of its letters); some are used by certain typefaces or styles of handwriting. This paper discusses these several kinds of variation among basic shapes, and gives examples from several scripts.

Within many scripts and written languages, there are certain letters which can each be embodied<sup>1</sup> by marks with two or more basic shapes. For each of these letters the language or its script normally has rules which prescribe the conditions under which each of those basic shapes is used to realize it. The present paper will discuss four kinds of such conditioned variation: positional variation, dialectal variation, variation among coscripts, and stylistic variation. For each of these, it will show examples of the alternative characters involved and will describe how the choice is made among them.<sup>2</sup>

## *Positional variations*

Positional variation occurs when a letter is realized by different basic shapes according to the place where it occurs within a syllable, a word, or another definable unit of its written language. Figures 1 through 3 give examples of such variation in three scripts.

In the Naskhi script of the Araboid genus (Fig. 1),<sup>3</sup> as in the other scripts of that genus, almost every letter can be realized by four different basic shapes which are often, but need not be, much alike. (Some letters which can never be attached to a following

Figure 1. Positional variation: *Naskhi script*.

Variants for word-initial position	Variants for word-medial position	Variants for word-final position	Variants for isolated position
ث	ث	ث	ث
ح	ح	ح	ح

Figure 2. Positional variation: *Neohellenic script*.

Variant for word-final position if not capitalized	Variant for all other positions if not capitalized
ς	σ

Figure 3. Positional variation: *Bengali script*.

Variants for syllables not containing consonants	Variants for syllables containing consonants
ৱ	ৱ
ৱ	ৱ

letter have only two basic shapes each.) The choice among these basic shapes depends on whether the letter occurs initially, medially, or finally in a word or occurs in isolation unattached to any other letter.

In the Neohellenic script (Fig. 2) the letter named “sigma,” if it is not capitalized, can be realized by either of two basic shapes. One of these occurs at the ends of words, the other occurs everywhere else.

In the Bengali script (Fig. 3), as generally in scripts of the Brahmic family, each vowel letter can be realized by two different basic shapes. One of these is used when the vowel forms an entire syllable by itself; the other is used when it occurs in a syllable together with one or more consonants.

### *Dialectal variations*

A script or an alphabet, like any other part of a language, may show dialectal variation. And—just as a language can be one language despite geographical, social, or other kinds of variation in the sounds which embody its phonemes, or variation in the grammatical constructions which it allows, or variation in the words which it uses for certain meanings—so a language or a script can be one language or one script despite dialectal variations in the marks which embody its letters. The status of any linguistic variety as a language or as a dialect depends only on usage, including popular opinion and official recognition. It cannot be determined by quantitative measurements, since the differences among the acknowledged dialects in one linguistic situation may be larger than the differences among the acknowledged languages in another situation.<sup>4</sup> Therefore, even though a writing system has some variation within it, it is one script with dialects so long as the people who use it regard it as a single script.

An example of dialectal variation on a largely geographical basis occurs within the Devanagari script. Although most letters of that script are printed by the same basic shapes wherever that script is used, some of its letters are printed by different basic shapes, depending on where the printing is done. Examples of these differences are shown in Figure 4; the two groups of characters

Figure 4. Dialectal variation: *Devanagari script*.

Hindi dialect  
variants

Bombay dialect  
variants

अ

अ

ओ

ओ

ऋ

ऋ

ल

ल

Figure 5. Compound script: coscripts and sets of basic shapes: *Neorman script*.

Majuscule set  
of basic shapes

Minuscule  
set of  
basic shapes

Italic  
set of  
basic shapes

Capital  
coscript

Italic  
capital  
coscript

Small  
capital  
coscript

Lower-case  
coscript

Italic  
lower-case  
coscript

A

*A*

A

a

*a*

D

*D*

D

d

*d*

G

*G*

G

g

*g*

N

*N*

N

n

*n*

S

*S*

S

s

*s*

V

*V*

V

v

*v*

shown there seem to have no generally accepted names, and the terms used for them there are taken from Lambert (1953). Within the region where the Devanagari script is used, types with basic shapes from the Hindi dialect seem to be used more commonly by printers in the north and east, while types with basic shapes from the Bombay dialect seem to be used more commonly by printers in the south and west. Many printers, however, use types with some basic shapes from each dialect. The present state of these dialects deserves systematic, scientific investigation.

Dialect differences within a script may extend to the languages which have alphabets from that script. Thus, the Hindi language is used in a region which extends across the boundary between these two dialects of the Devanagari script, and the Hindi alphabet from that script has dialects which correspond to the dialects of its script. Attempts to eliminate these dialectal differences and to create a single compromise alphabet for Hindi were made at two conferences organized by the Uttar Pradesh government during the 1950's (Saran 1969, pp. 61-62).

#### *Variations among coscripts*

Each of the letters of the Neoroman script can be embodied by several different characters, which may differ from one another in basic shape, in height, and/or in slope. These characters can be divided into groups, each of which contains at least one character for each letter of the script, and each of which is used under certain definable circumstances as the source for characters to embody the letters of the script. Therefore, when a person writes with an alphabet from the Neoroman script, he must consider not only the letters which represent the words he wants to write but also the groups of characters that are appropriate for those words or parts of those words. Each character which he writes must belong to both the appropriate letter and the appropriate group of characters.

Because each of these groups of characters within the Neoroman script could itself be a script—except that it is not normally used alone to provide the writing system of any language—each of these groups may be called a “coscript” within the Neoroman script, and the Neoroman script as a whole may be called a “compound

script.”<sup>5</sup> The Neoroman script ordinarily has five coscripts: “capitals,” “italic capitals,” “small capitals,” “lower-case,” and “italic lower-case.”<sup>6</sup> Examples of them are shown in Figure 5.

The characters belonging to the different coscripts of a compound script need not be wholly dissimilar, but they may show various kinds of similarity. In Figure 5 the five Neoroman coscripts have been grouped according to the sets of basic shapes which they use. One set of basic shapes, which may be called the “minuscule” set, is used only by the lower-case coscript; another, which may be called the “italic” set, is used only by the italic lower-case coscript. The capital, italic capital, and small-capital coscripts, however, all use the same set of basic shapes, which may be called the “majuscule” set.<sup>7</sup> These three sets of basic shapes are non-discrete, because some of their basic shapes belong to more than one set.

Some other compound scripts are illustrated in Figures 6 through 10. Each of these illustrations shows only a few letters from its script, but should suffice to show the kinds of variations which occur among its coscripts and sets of basic shapes. In each of these illustrations, the characters shown on one horizontal row embody the same letter.

The Fraktur script (Fig. 6) has capital and lower-case coscripts; the sets of basic shapes which its coscripts use are essentially discrete.<sup>8</sup> The Gaelic script (Fig. 7) also has capital and lower-case coscripts, but the sets of basic shapes which they use are non-discrete. The Neocyrillic script (Fig. 8) has five coscripts and three sets of basic shapes, corresponding to those of the Neoroman script.<sup>9</sup>

Although compound scripts seem to be most common within the Hellenic family, at least two languages of the Far East use compound scripts belonging to other families. The Cambodian script (Fig. 9) has four coscripts and three non-discrete sets of basic shapes.<sup>10</sup> The script of the Japanese syllabary (Kana script, Fig. 10) has two coscripts which use separate, discrete sets of basic shapes.<sup>11</sup>

Figure 6. Compound script: coscripts and sets of basic shapes: *Fraktur script*.

Majuscule set of  
basic shapes

Capital coscript

B

G

S

R

Minuscule set of  
basic shapes

Lower-case coscript

b

e

h

r

Figure 7. Compound script: coscripts and sets of basic shapes: *Gaelic script*.

Majuscule set of  
basic shapes

Capital coscript

A

F

S

T

Minuscule set of  
basic shapes

Lower-case coscript

a

f

r

t

### *Stylistic variations*

In linguistics and literary scholarship “style” has been a word of many meanings. As it is being used here, a style is a choice: any choice which a language allows a speaker or a writer to make among several sets of linguistic forms that have the same denotative meaning. In written texts printed from moveable type or by other mechanical devices for reproducing characters, one such stylistic choice is the choice among typefaces. Although the selection of a typeface for printing a certain text usually has no effect on the basic shapes which realize the letters in that text, it sometimes happens that a letter, even though its coscript, its position, and its dialect are known, may still be realized by any of two or more basic shapes, the selection among which depends on the typeface which has been chosen. Figures 11 and 12 show examples

Figure 8. Compound script: coscripts and sets of basic shapes: *Neocyrillic script*.

Majuscule set of basic shapes		Minuscule set of basic shapes		
Capital coscript	Italic capital coscript	Small capital coscript	Lower-case coscript	Italic lower-case coscript
А	<i>А</i>	А	а	<i>а</i>
Г	<i>Г</i>	Г	Г	<i>г</i>
Е	<i>Е</i>	Е	е	<i>е</i>
И	<i>И</i>	И	и	<i>и</i>
К	<i>К</i>	К	к	<i>к</i>
У	<i>У</i>	У	у	<i>у</i>

Figure 9. Compound script: coscripts and sets of basic shapes: *Cambodian script*.





















Slanted set of basic shapes		Round set of basic shapes	Cambodian set of basic shapes
Slanted coscript	Standing coscript	Round coscript	Cambodian coscript
			
			
			
			
			

Figure 10. Compound script: coscripts and sets of basic shapes: *Kana script*.

Katakana set of basic shapes	Hiragana set of basic shapes
Katakana coscript	Hiragana coscript
	
	
	
	
	
	

of such stylistic variations among basic shapes. The two shapes from the Neoroman script (Fig. 11) may both be used to realize the letter g when it is not capitalized; the two shapes from the Neocyrillic script (Fig. 12) may both be used to realize the letter П. In each script, some typefaces use the first mark, while others use the second.

A similar variation among basic shapes occurs in the handwriting of some people. The writer of the example shown in Figure 13 (which is taken from a marginal notation on a manuscript) has used two basic shapes for non-capitalized r and two for non-capitalized s. The variation between the shapes for r must be a stylistic variation, since both occur in equivalent environments. The data shown here is insufficient to prove whether the shapes for s are stylistic or positional variants.

*A comment on the linguistic treatment of these variations*

The four kinds of variation which have been described here should be dealt with in the linguistic analysis of any script which has them. The variation resulting from a choice among typefaces or styles of handwriting is, as has been said, a stylistic variation and can be described by whatever notation is used to handle other stylistic variations. Dialectal variation is a similar phenomenon, and presumably can be handled by a similar notation. The other two kinds of variation are both conditioned by the contexts in which letters occur. For positional variation, the context is formed by other letters (or equivalent elements such as the space between words). For variation among coscripts, the context is formed by certain linguistic elements that occur simultaneously with letters and specify which coscript has been chosen for them.

1. The relationship between abstract units of language (such as letters and basic shapes) and the objects which physically serve as a communication medium for a language will be called "embodiment," while the relationship between two kinds of abstract linguistic units will be called "realization." Thus, a mark embodies a letter or a basic shape, while a basic shape realizes a letter.

Figure 11. Stylistic variation: *Neoroman script*.

g g

Figure 12. Stylistic variation: *Neocyrillic script*.

П ^

(note that  
in handwritten  
ones, however,  
they may really  
be in free vari-  
ation — not  
mere stylistic  
variants.)

Figure 13.  
Example of stylistic variation:  
“*Neoroman*” handwriting.

2. The author wants to thank Norman H. Zide and George L. Trager for their comments and assistance.
3. The system of names used here for scripts and genera of scripts was defined in Herrick 1974.
4. If one linguistic situation is described by both terms, however, its languages will be more divergent than its dialects. For the linguistic use of these terms, see Hockett (1958, pp. 321-22), Gleason (1961, p. 398), and Trager (1972, pp. 12-14).
5. A script which does not have coscripts may be called a "simple script."
6. Depending on the linguistic analysis being used, the Neoroman script may also have two or four boldface coscripts.
7. The term "alphabet" has been used both for coscripts and for sets of basic shapes. Gill (1951, pp. 257-61), for example, uses it to mean sets of basic shapes, while Carter (1969, p. 45) refers to all the Neoroman coscripts as "alphabets," and then refers to the three oldest coscripts, which use different sets of basic shapes, as the "essential alphabets" of that script.
8. There is one Fraktur letter which is always realized by the same basic shape, whether it is capitalized or not.
9. See Trager (1972, p. 261) on the use of Neocyrillic small capitals.
10. See Huffman (1970, pp. 77-79) for the appearance and uses of these Cambodian coscripts.
11. See Chaplin and Martin (1967, pp. 6-9) for the contrasting uses of these Japanese coscripts.

## REFERENCES

- Carter, H. *A view of early typography: up to about 1600*. Oxford: Clarendon Press, 1969.
- Chaplin, Hamako I., and S. E. Martin. *Manual of Japanese writing*. Vol. I. New Haven: Yale University Press, 1967.
- Gill, E. Typography. In P. A. Bennett (ed.), *Books and printing: a treasury for typophiles*. Cleveland: World Publishing Co., 1951.
- Gleason, H. A., Jr. *An introduction to descriptive linguistics*. (Rev. ed.) New York: Holt, Rinehart & Winston, 1961.
- Herrick, E. M. A taxonomy of alphabets and scripts. *Visible language*, 1974, 8, 5-32.
- Hockett, C. F. *A course in modern linguistics*. New York: Macmillan, 1958.
- Huffman, F. E. *Cambodian system of writing and beginning reader*. New Haven: Yale University Press, 1970.
- Lambert, H. M. *Introduction to the Devanagari script: for students of Sanskrit and Hindi*. London: Oxford University Press, 1953.
- Saran, S. *India's national writing*. Delhi: Munshiram Manoharlal, 1969.
- Trager, G. L. *Language and languages*. San Francisco: Chandler Publishing Co., 1972.