

A Taxonomy of Alphabets and Scripts

Earl M. Herrick

This paper describes a taxonomy or system of classification to describe alphabets and to show the various degrees of formal similarity which they bear to one another. The principles of this taxonomy, largely borrowed from the biological sciences, are briefly stated; three taxonomic levels, corresponding to three degrees of similarity, are then defined. For each level, the paper describes the kinds and degrees of similarity which alphabets must have to be included within one taxon; these similarities are illustrated by several different kinds of taxa and some of the alphabets which belong to them. Several problems in the comparison of writing systems which will require further research are mentioned.

Linguists, typographers, and others who work with written language do not presently have any adequate system of classification for describing similarities and differences among the alphabets and scripts which are used to write languages. This paper will describe one method by which such a taxonomy or classification system may be constructed, and it will offer an outline of such a taxonomy.

A linguistic taxonomy of this kind should be helpful in several ways to all those who study and work with written languages.¹ It will allow the description of any alphabet—more or less accurately, as may be desired—by naming the script or another of the taxa (i.e., groupings in the taxonomy) to which it belongs. It will also allow the similarity between any two alphabets to be described by naming the lowest taxon (i.e., the smallest grouping) to which they both belong. The writing of a taxonomy such as this will also help to develop terminology which may be useful in the further study of written language.

1. The terms “written language,” “writing,” and “written form” of a language, as they are used in this paper, refer to written rather than to spoken language and to printing rather than to handwriting. In terms of the distinctions made by Mountford (1968), this paper deals with “writing-1” but does not deal with “writing-3.”

Because many descriptions of written languages have already been published, it may seem strange that no scientific taxonomy of alphabets and scripts has yet been constructed. It happens, however, that all previous treatments of written language have been directed to other purposes.

Some works, such as those of Diringer (1968) and Jensen (1969), are filled with examples of texts and scripts intended to show the historical development of written language; unfortunately, they have not organized their examples in any systematic, scientific manner. Other works, like that of Giljarevskij and Grivnin (1964), list alphabets from many parts of the world, but do not systematically describe the internal organization of these alphabets, their relationships to one another, or their relationships to the rest of their respective languages. Trager (1972) has developed a system of linguistic terminology for describing various kinds of relationships between speech and writing. The many illustrations of alphabets and other writing systems which he includes are generally arranged to show the similarities and relationships among them. Gelb (1963) has developed a general linguistic theory of written language which will surely serve as the basis for all future scientific work in this field. Gelb's interest, however, is in theory rather than in description, and he does not try to describe, except incidentally, the structure and functioning of individual present-day writing systems.

There are also certain works which are ostensibly general classifications of alphabets, but which are seen on closer inspection to be only classifications of roman typefaces. An example is the *Klassifikation der Schriften* published by the Deutsche Industrie Normenausschuss (DIN 16 518).² This taxonomy starts out by lumping together all non-roman writing, and then proceeds to describe and subclassify roman typefaces. Although this taxonomy might be used to describe these roman typefaces³ after that category has been defined by a more inclusive taxonomy,⁴ it cannot serve as a general taxonomy of alphabets.

2. See Noordzij (1970) for a description of this classification.

3. Although Noordzij casts doubt on its usefulness even for this purpose.

4. Previous writers on scripts and alphabets have often failed to agree on just what the roman script or alphabet is: see Tanselle (1967) on the problem of putting the

Certain other descriptions of alphabets are typified by the work of Downie (1963). This is essentially a typefounder's specimen book. It is designed to tell a printer the languages which can be printed with the types available from its publisher, it gives him the catalog information which he needs in order to buy fonts for each language, and (for some languages) it warns him of unusual letters for which he must have types or matrices before printing in them. Although Downie supplies short commentaries on each script for which Monotype makes matrices, his remarks are sometimes tantalizingly incomplete.⁵ And although he mentions that certain alphabets are related to certain others, he does not try to describe their relationships systematically.

Principles for a Linguistic Taxonomy

A scientific taxonomy such as this one should have five characteristics.

1. The taxonomy should be *useful*. On the need for this, John Stuart Mill has written, "There is no property of objects which may not be taken, if we please, as the foundation for a classification or mental grouping of those objects. . . . [However,] the ends of scientific classification are answered when the objects are formed into groups respecting which a greater number of general propositions can be made, and those propositions more important, than could be made respecting any other groups into which the same things could be distributed. . . . The test of [the] scientific character [of a classification] is the number and importance of the properties which can be asserted in common of all objects included in the group" (Mill, 1881, pp. 498–499).
2. The taxonomy should be *exhaustive*. It should attempt to treat with equal precision all branches of the field which it claims to classify—

term "italic" into the same taxonomy as the terms "roman" and "latin"; Noordzij (1970) on the relationship between fraktur and roman; Gill (1951), who avers that English uses three roman alphabets, some of which share certain letter shapes; and Hoch (1968), who says that English uses "at least two if not three distinct roman alphabets together." The author has therefore avoided the term "roman" in the present paper, and has defined and used instead the terms "neoroman" and "romanoid."

5. See, for example, his discussion of Hebrew vowel points (pp. 7–8).

the smaller branches along with the larger, the obscure branches along with the more familiar. Some of these branches may be more complex than others and may require finer subclassification. Other branches may be poorly known, and the taxonomist may have to say that his work on them is tentative and that their satisfactory treatment awaits further research. But a taxonomist should not claim to have classified an entire field when he has, in fact, classified only a part of it.

3. The taxonomy should be *hierarchical* (or *graduated*, as Tanselle [1967] has used that term). Alphabets, like all parts of language, are immensely variable, and two alphabets may show any degree of similarity from virtual identity to utter dissimilarity. In order to be useful, a taxonomy of alphabets should therefore be able to describe many different degrees of similarity. A one-level taxonomy, in which two different alphabets either do or do not belong to the same taxon, cannot do this. The taxonomy must have several levels, so that alphabets which are more closely related can be placed together in a lower taxon, while those which are less closely related can be placed together only in a higher taxon.

4. The taxonomy should be *precise*. Its several taxa should be described so well that any individual—in the present taxonomy, any alphabet—can be placed in its proper taxa.

5. The taxonomy should be *tolerant* of unforeseen variation. Although a taxonomy such as this must be precise, it cannot achieve its precision by using rigidly defined criteria. Systems of writing—like all linguistic phenomena—are too diverse to be fitted into any absolute system of categories. Even if a linguist were to devise a system of rigid criteria which would successfully classify all the writing systems that he knew, he would have no reason to believe that the next writing system which he met could be classified successfully into the same categories by the same criteria. For a taxonomy to be both tolerant of such unforeseen variation and precise enough to be useful, it must be based not on rigid criteria but on the considered judgement of those who have developed it and are familiar with its material. A taxonomy such as this should therefore simply state that certain alphabets or certain scripts are similar enough to form a taxon of a certain level. Any other script which is as much like the scripts of one taxon as they are like one another will also be placed in that taxon.

Scripts which show a like degree of similarity to one another but not to those of a taxon already established will be placed in another taxon of the same level. In this kind of taxonomy, therefore, the similarity between scripts which is necessary for membership in the same taxon should not be a stated similarity on certain stated points, but should be a degree of general similarity implied by the mutual similarities of those alphabets and scripts which are already classified together into taxa.

Linguistics is not alone in its need to describe phenomena which show great, unforeseeable variation. It shares this need with all the other sciences which have to do with living things—the biological sciences and the social sciences. In this respect, these sciences are unlike the physical sciences in which, for example, Mendeleev's classification of the chemical elements has its immense value precisely because it is so very rigid. The author has therefore borrowed for this taxonomy of alphabets and scripts several ideas from botanical and zoological taxonomies. Those taxonomies have problems and uncertainties which are much like those of the taxonomy presented here. But biology, which has been evolving its modern taxonomies for more than two centuries, has found useful solutions for some of these problems; and even its unsettled problems have led to discussion which may help clarify similar problems in linguistic taxonomies.⁶

Alphabets and Scripts

One distinction which must be made by any adequate description of written language is the distinction between an alphabet and the

6. One difficulty which linguistic and biological taxonomists share is an occasional difficulty in defining the basic groups of individuals which they should classify. For linguists these groups are speakers of "languages"; for biologists they are members of "species." These groups of individuals are often defined in roughly equivalent ways—languages by the possibility of mutual understanding among neighboring speakers; species by the possibility of fertile interbreeding among neighboring individuals. But compare Hockett (1958, pp. 323–325) on the linguistic difficulty of distinguishing the French and Italian languages because the inhabitants of some Alpine villages can understand the undoubtedly French and undoubtedly Italian speech of their neighbors on either side, and de Laubenfels (1952) on the biological difficulty of distinguishing yellow-bellied and red-bellied sapsuckers because these two kinds of birds interbreed and hybridize in a small area in Oregon.

script to which it belongs. An alphabet is part of one language; an alphabet has a definite, complex internal structure; an alphabet has close relationships with certain other parts of its language. On the other hand, a script does not necessarily belong to any one language; a script has much less internal structure than an alphabet; a script has no direct relationships with the other parts of any language.

The *alphabet* of a language is made up of graphemes, which are abstract linguistic units of written language. Most of the graphemes of an alphabet are letters, but an alphabet may also have supra-segmental graphemes—e.g., the graphemes of capitalization or the graphemes of stress, such as the diacritic grapheme “accent” in Spanish and in Czech.

Each letter of an alphabet has five characteristics.

1. A letter has a name. This means that whenever a written mark is an embodiment of a letter of a certain language, any normal literate of that language can name the letter which it embodies.
2. A letter has a pronunciation; i.e., it has certain correspondences with linguistic units in the spoken form of its language. The complexity of these correspondences varies from one language to another and from one letter to another within a language. For example, the letters of Finnish all have simple, regular correspondences with the phonemes of spoken Finnish, while many letters of English and Thai have awkward or irregular correspondences with the phonemes of their respective languages (Hanna, et al., 1966; Haas, 1956).
3. A letter has a place in the alphabetical order of its language. This means that normal literates who are given a group of written words (or other strings of characters that embody letters) can arrange them according to this alphabetical order.
4. A letter, and every other grapheme, has a certain role in the graphotactics of its language. Each language has graphotactic formulas which state the combinations of graphemes which may occur in that language. The graphotactic formulas of a language are peculiar to it and are sometimes highly characteristic of it; for example, Polish words may contain the letter sequence *szcz*, but English words do not contain that sequence unless they are unassimilated borrowings from Polish.⁷

5. A letter is embodied by marks which have a certain basic shape or a certain few basic shapes. The basic shape of a character is itself an abstract linguistic unit; it is the group of geometrical distinctive features which a written mark must have so that a literate person will recognize it as an embodiment of a certain letter.⁸ These basic shapes are what Goudy has called the “fundamental forms” (1940, pp. 28–29) or “intrinsic shapes” (1942, pp. 86, 90) of letters. It is to such a basic shape, and the way it remains invariant under stylistic and suprasegmental changes, that Gill (1951) referred when he said that “a Roman Capital A does not cease to be a Roman Capital A because it is sloped backwards or forwards, because it is made thicker or thinner, or because serifs are added or omitted.”

When a letter can be represented by more than one basic shape, the choice of shape to be used may depend on the suprasegmental graphemes that are being represented along with the letter. For example, the English letter a uses the basic shape **A** when it is embodied together with either of the suprasegmental graphemes “capitals” or “small capitals.” If it is not capitalized, it may be represented by either of the basic shapes **a** or **ɑ**, but the latter shape is considered typical of the suprasegmental grapheme “italics.” In many languages there are letter graphemes which have alternative basic shapes conditioned by other segmental graphemes.⁹ For example, the modern Greek letter Σ (if it is not capitalized) is represented by the basic shape σ when it occurs at the end of a word and by the basic shape σ when it occurs in any other position.

7. In the graphotactic formulas, in the phoneme-grapheme correspondences, and in the alphabetical order of a language the letters of its alphabet usually act as separate units. There are, however, many exceptions to this generalization, and a thorough comparative study of written languages would have to deal with cases in which several letters act together as single units for some of these purposes.

8. In Herrick (*forthcoming*) a basic shape is called a “graphemic sign” and a geometrical distinctive feature is called a “graphon.”

9. The f-ligatures of English (ff fi fl ffi fll) are embodiments of neighboring letters which condition each other’s embodiment. For lack of space, the present paper has not thoroughly described the ligatures for each language, although the way in which the alphabets of a taxon form their ligatures is often a point of similarity among them.

A *script*, like an alphabet, is made up of graphemes. Most of its graphemes, like those of an alphabet, are letters, although it may also have suprasegmental graphemes. A script has graphotactic formulas which specify merely the ways in which its letters and suprasegmentals may co-occur. For each set of co-occurring graphemes the script specifies the basic shape and the other features of the character which will embody them. A script has no other internal organization or external relationships of its own. It has no tactic formulas to specify how its letters may follow one another. Its letters have no pronunciations, and they have no alphabetical order.

It happens, however, that in the minds of many people some letters of a script may have expected pronunciations and expected orderings. These expectations are suggested to them by the alphabets which they know and which belong to the script. For example, in the script to which the English alphabet belongs, the letter a will often be thought of as the first letter of the script, and it will often be thought of as corresponding to a low vowel phoneme, because these facts are true of the letter a in most of the alphabets of this script.

Criteria for This Taxonomy

Three kinds of alphabetical similarity have been used as the criteria for constructing this taxonomy:

1. Similarity in the basic shapes of letters.
2. Similarity in the correspondences between letters and phonemes.
3. Similarity in the alphabetical orders of letters.

There are other ways in which written languages may be similar. Some of these kinds of similarity would not be satisfactory criteria for any such taxonomy because they refer to features that tend to be less permanently characteristic of individual written languages. Others could not be integrated into this taxonomy because they would group alphabets in ways that would cut across the groupings produced by the criteria already being used. Three kinds of similarity which have not been used as criteria for this taxonomy but which should be treated by more general studies of written language are:

1. The directions of writing within a text.
2. The extent to which phonemes of various kinds are represented by letters.

3. The relative locations of the letters representing various kinds of phonemes. (One method for positioning vowel and consonant letters is described in the appendix to this paper.)

Scope of This Taxonomy

For a language which is both spoken and written, the graphemes of its written form can represent various kinds of linguistic units in its spoken form. The graphemes of a language which correspond to its vowel and consonant phonemes may be called the letters of its alphabet. Graphemes which correspond to words or to linguistic units of the semantics may be called logograms.

The taxonomy of alphabets and scripts described here is not intended to include logograms. It will, however, include syllabaries. The distinction between alphabets and syllabaries is awkward to draw in theory and harder to apply in practice, since the graphemes of both correspond to the phonemes of spoken language. It has sometimes been said that the letters of an alphabet represent all the phonemes of a language, while the marks of a syllabary correspond to entire syllables, or correspond to only the consonants of a language, or correspond to the consonants more completely than to the vowels. (This variety of definitions should give the reader some warning of the problems involved here.) Unfortunately for the writers who make this distinction, there is no language in which the writing system represents every phoneme of the spoken language. The writers who make this distinction disagree, moreover, on the precise size, separateness, location, and scarcity of vowel letters which are required to demote an alphabet to a syllabary.¹⁰ A complete theory of written and spoken language would have to make several interrelated distinctions in this matter, but a simple dichotomy of alphabets and syllabaries is not sufficient.

10. The reader should be warned that some writers have used the term "syllabary" as a pejorative. Historically, syllabaries preceded alphabets. A writer may choose to infer from this that a people which writes with a syllabary is more primitive (whatever that may mean) than a people which writes with an alphabet; and he may then try to show that his own language uses an alphabet while that of some people which he wants to disparage uses a syllabary. This whole game is a messy business which the present author does not want to get into.

Figure 1. Sketch of this sample taxonomy.

<i>Alphabets</i>	<i>Scripts</i>	<i>Genera of Scripts</i>	<i>Families of Scripts</i>
Czech	} Neoroman	} Romanoid*	} Hellenic
Dinka			
English			
German			
Hawaiian			
Icelandic			
Irish			
Kazakh			
Latin			
Malay			
Navajo			
Serbo-			
Croatian			
Spanish			
Swedish			
Turkish			
Latin	} Paleoroman Fraktur Gaelic	} Hellenic	
German			
Irish			
Buryat	} Neocyrillic*	} Cyrilloid	
Chuvash			
Kazakh			
Russian			
Serbo-			
Croatian			
Udmurt			
Ukrainian			
Yakut			
Old Church			
Slavic	} Paleocyrillic	} Hellenoid	
Modern Greek			
Coptic			

NOTES

1. Languages which are named more than once in this taxonomy have been written by alphabets of more than one script.

2. Many of these taxa have members other than those shown here.

*Taxon illustrated in this paper.

— — — Taxon for which no name is proposed here.

A Sample Taxonomy of Alphabets and Scripts

Kinds of taxa. The taxonomy of alphabets and scripts which is suggested by this paper has three regular levels of taxa higher and more inclusive than the individual alphabet. In order of increasing size and decreasing similarity, these are called *scripts*, *genera* of scripts (singular *genus*), and *families* of scripts.

The degree of similarity among alphabets or scripts which is necessary to place them in the same taxon on each of these levels is defined by the taxa which are described and illustrated here. The kind of similarity which is necessary to place alphabets or scripts in the same taxon varies somewhat from level to level of the taxonomy.

For alphabets to be included in the same script, their letters must have almost the same alphabetical orders and almost the same pronunciations, and they must be embodied almost entirely by characters with identical basic shapes. Characters from two languages have identical basic shapes if a person literate in both languages who sees the characters in isolation cannot tell which came from which language. From a printer's point of view, letters from two languages are embodied by identical basic shapes if he can use the same type to print either letter without causing a reader to think that something unusual has occurred.

For scripts to be included in the same genus, or for genera to be included in the same family, the letters of their alphabets must have similar alphabetical orders and similar pronunciations, and the basic shapes which embody them must show certain patterns of similarity from one smaller taxon to another within the larger taxon. Although the same criteria apply to both of these taxonomic levels, much greater similarity is expected among alphabets of the same genus than is expected among alphabets of the same family but different genera.

An outline of this taxonomy. Figure 1 contains a sketch of the taxonomy of alphabets and scripts which is suggested by this paper.¹¹ The author has availed himself of the taxonomist's traditional prerogative to make up names for the taxa which he is the first to propose. The names

11. The taxonomy presented here has been developed from the author's earlier classification for alphabets (Herrick 1966, pp. 13-23). All of his acknowledgements should, of course, be repeated here. The author also wants to thank P. K. Saha, Robert A. Palmatier, and Hardy Carroll for their helpful comments on the present paper.

given to the genera of scripts all have the suffix “-oid”; these names should be distinctively marked, since the genera of scripts in a completed taxonomy will be few enough that anyone interested in the study of written language can learn something about all of them, and therefore those names will be frequently used to describe alphabets and scripts. In a few cases where a name has not been made up for a taxon, a line of dashes appears in Figure 1.

Scripts. The similarity between alphabets which is necessary for their inclusion in the same script is suggested by the alphabets and scripts named in Figure 1 and is illustrated in Figures 2 and 3. Each of these scripts is exemplified by two or more of its alphabets. The listing for each alphabet shows its letters in their alphabetical order, the pronunciations of its letters, and the basic shapes which embody them. These pronunciations are shown here by characters of the International Phonetic Alphabet¹² written between square brackets; no more than three pronunciations are shown for each letter. Phonetic symbols listed in this way can only suggest the actual correspondences between the letters and phonemes of a language, but they will serve well enough to show the similarities in pronunciation which must exist among the alphabets of a script. If two or more basic shapes are shown for a letter, they may be stylistic variants, variants conditioned by segmental or suprasegmental graphemes, or variants corresponding to different pronunciations; no more than three basic shapes are shown for each letter. In Figures 2 and 3 the letters which are embodied by the same basic shapes have been placed together, insofar as this may be done without altering their respective alphabetical orders.

The Neocyrillic script (Fig. 2) is probably the world's second largest script—next in size only to the Neoroman script—and is now used for a large number of languages which occur in the Soviet Union and in several other countries. It is exemplified here by three of its alphabets. The name “Neocyrillic” has been used to distinguish these alphabets from alphabets like that used for the Russian language before Tsar

12. The International Phonetic Alphabet may be conveniently found in recent editions of *Encyclopædia Britannica* in the article “Phonetics,” as well as in many linguistic textbooks.

Russian Alphabet			Ukrainian Alphabet			Serbian Alphabet		
Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes
А	[a,ə]	а а	А	[a]	а а	А	[a]	а а
Б	[b]	б б	Б	[b]	б б	Б	[b]	б б
В	[v]	в в	В	[v,w]	в в	В	[v]	в в
Г	[g]	г г	Г	[h]	г г	Г	[g]	г г
Д	[d]	д д	Д	[d]	д д	Д	[d]	д д
Е	[e,je,jo]	е е	Е	[æ]	е е	Е	[e]	е е
Ж	[ʒ]	ж ж	Є	[jɛ]		Ж	[ʒ]	ж ж
З	[z]		Ж	[ʒ]		З	[z]	
И	[i+i,j]	и й	И	[ɛ]	и	И	[i]	и
			І	[i]	і			
			Ї	[ji]	ї			
			Й	[j]	й			
К	[k]		К	[k]		Ј	[j]	ј
Л	[l]	л л	Л	[l]	л л	К	[k]	
М	[m]		М	[m]		Л	[l]	л л
Н	[n]		Н	[n]		Љ	[ʎ]	љ
О	[a,o,ə]		О	[o]		М	[m]	
П	[p]	п п	П	[p]		Н	[n]	
Р	[r]	р р	Р	[r]		Њ	[ɲ]	
С	[s]		С	[s]		О	[o]	

Figure 2. Alphabets of the Neocyrillic script.

Peter I's reform of the Russian writing system.¹³ The former Russian alphabet and the alphabet still used to print texts in the Old Church Slavonic language would belong to the Paleocyrillic script.

The Bengali script (Fig. 3) includes the Bengali and Assamese alphabets, which are shown to exemplify it, and the alphabets of a

13. Tsar Peter I's reform of the Russian writing system has been described by Kaldor (1969-70).

تنوع القاطن السبع الطوال

KUḤFĪ

اذ قال رجل يا رسول الله نمرة فلبسها
فوما بها الى الرجل صنعت

NASKHĪ

الرحمون برحمهم الرحمن رحمة في الارض برحمتهم في السماء
الرفعة في خط الرفعة

RUQḤAH

الجمال سحر يظفي على الكون روعته وتسير في موكبه الضنون
نتعليق أو الفارسي

NASTAʿLĪQ OR FĀRISĪ

الجمال سحر يظفي على الكون روعته وتسير في موكبه الضنون

DĪWĀNĪ

Figure 4. Writing in several varieties and alphabets of the genus Araboid.

there are certain seldom-used styles of printer's type, such as Civilité, which probably constitute other scripts of the genus Romanoid. The Palcoroman script includes only the alphabet of classic Latin, which had no suprasegmental graphemes and only one basic shape for each letter.

The genus Telugoid (Fig. 6) is exemplified by its only two alphabets, each of which constitutes a separate script. The Kannada and Telugu languages, like many other languages of India, are examples of the

Neoroman Script			Fraktur Script			Gaelic Script		
Represented by the English alphabet			Represented by the German alphabet of that script			Represented by the Irish alphabet of that script		
Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes
A	[æ,ej]	a a	Ɑ	[a,ɛ]	Ɑ ä	Ɑ	[ɔ]	Ɑ
B	[b]	b	Ɱ	[b]	Ɱ	Ɱ	[b]	Ɱ
C	[k,s]		Ɱ	[k]	c	Ɱ	[k]	
D	[d]	d	Ɱ	[d]	Ɱ	Ɱ	[d,ð]	Ɱ
E	[ɛ,i]	e e	Ɱ	[e,ɛ,ə]	e	Ɱ	[e,ɛ]	e
F	[f]	f f	Ɱ	[f]	f	Ɱ	[f]	
G	[g,dʒ]	g g	Ɱ	[g]	g	Ɱ	[g]	
H	[h]	h	Ɱ	[h]	Ɱ	Ɱ	[h]	h
I	[i,aj]	i i	Ɱ	[i]	i	Ɱ	[i]	
J	[dʒ]	j	Ɱ	[j]	j			
K	[k]	k k̂	Ɱ	[k]	k̂			
L	[l]	l l̂	Ɱ	[l]	l̂	Ɱ	[l]	l̂
M	[m]	m	Ɱ	[m]	m	Ɱ	[m]	m
N	[n]	n	Ɱ	[n]	n	Ɱ	[n]	n
O	[o,ow,ə]		Ɱ	[o,ø]	o ö	Ɱ	[o]	
P	[p]	p	Ɱ	[p]	p̂	Ɱ	[p]	
Q	[k]	q	Ɱ	[k]	q			
R	[r]	r	Ɱ	[r]	r	Ɱ	[r]	r̂
S	[s,z]		Ɱ	[s,z]	ŝ f̂	Ɱ	[s,ʃ]	r̂
T	[t]	t t̂	Ɱ	[t]	t̂	Ɱ	[t,θ]	
U	[u,yuw,ə]	u	Ɱ	[u,y]	u ü	Ɱ	[u]	
V	[v]	v	Ɱ	[f]	v			
W	[w]	w	Ɱ	[v]	w			
X	[ks]		Ɱ	[ks]	x̂			
Y	[j]	y ŷ	Ɱ	[j]	ŷ			
Z	[z]		Ɱ	[ts]	ẑ			

Figure 5. Scripts of the genus Romanoid.

Kannada Script		Represented by the Kannada alphabet	
Letter	Pronunciation	Other basic shapes	
ಅ	[ə]		
ಆ	[a:]		
ಇ	[i]		
ಀ	[i:]		
ಊ	[u]		
ಋ	[u:]		
ಋ	[ɾɨ]*		
ಌ	[e]		
಍	[e:]		
ಐ	[ə]		
ಋ	[o]		
ಠ	[o:]		
ಉ	[əw]		
ಕ	[k]		
ಖ	[kh]*		
ಗ	[g]		
ಘ	[gh]*		
ಙ	[ŋ, ts]		
ಚ	[tʃh]*		
ಛ	[dʒ, dz]		
ಜ	[dʒh]*		
ಟ	[ʈ]		
ಠ	[ʈh]		
ಡ	[ɖ]		
ಢ	[ɖh]		
ತ	[t]		
ಥ	[tʰ]		
ದ	[d]		
ಧ	[dʰ]		
ನ	[n]		
ಪ	[p]		
ಫ	[pʰ]		
ಬ	[b]		
ಭ	[bʰ]		
ಮ	[m]		
ಯ	[j]		
ರ	[r]		
ಲ	[l]		
ವ	[v]		
ಶ	[ʃ]		
ಷ	[ʃʰ]		
ಸ	[s]		
ಹ	[h]		
ಳ	[ʎ]		
ಱ	[ɻ]		
ಽ	[ʎ]		

Figure 6. Scripts of the genus Telugoid.

linguistic situation which Ferguson (1959) has called “diglossa.”¹⁴ These languages contain many words which have been borrowed from Sanskrit but which are used only in literary, religious, and poetic speech and writing. These borrowed Sanskrit words are not usually assimilated to the pronunciation and writing systems of the borrowing languages, but retain certain sounds and letters which do not occur elsewhere in the borrowing language. This fact poses a large problem for the linguistic description of Kannada, Telugu, and the other Dravidian languages of southern India. All of these languages tend to have many fewer consonant phonemes than Sanskrit and its related

14. Ferguson defines diglossa as “a relatively stable linguistic situation in which, in addition to the primary dialects of the language . . . there is a very divergent, highly codified . . . superposed variety, the vehicle of a large and respected body of written literature . . . which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.”

Genus Devanagaroid			Genus Telugoid			Genus Sinhaleoid		
Represented by the Hindi alphabet of the Devanagari script			Represented by the Telugu alphabet of the Telugu script			Represented by the Sinhalese alphabet of the Sinhalese script		
Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes	Letter	Pronunciation	Other basic shapes
अ	[ə]	अ	అ	[ə]	అ	අ	[ə]	අ
आ	[a]	आ	ఆ	[a:]	ఆ	ආ	[a]	ආ
इ	[i]	इ	ఇ	[i]	ఇ	ඇ	[æ]	ඇ
ई	[i:]	ई	ఈ	[i:]	ఈ	ඈ	[æ:]	ඈ
उ	[u]	उ	ఉ	[u]	ఉ	ඉ	[i]	ඉ
ऊ	[u:]	ऊ	ఊ	[u:]	ఊ	ඊ	[i:]	ඊ
ऋ	[r]	ऋ	ఋ	[r+]	ఋ	උ	[u]	උ
ॠ	[e]	ॠ	ౠ	[e]	ౠ	ඌ	[u:]	ඌ
ऌ	[æ,æj]	ऌ	ఌ	[e:]	ఌ	ඍ	[ε]	ඍ
ॡ	[o]	ॡ	఍	[ə]	఍	ඎ	[e:]	ඎ
ॢ	[ɔ,ɔw]	ॢ	ఊ	[o]	ఊ	ඏ	[ɔ]	ඏ
			ఋ	[o:]	ఋ	ඐ	[ow]	ඐ
			ౠ	[ow]	ౠ	එ	[aj]	එ
						ඒ	[aw]	ඒ
						ඓ	[ri]	ඓ
						ඔ	[ri:]	ඔ
						ක	[k]	ක
						ඛ	[kh]	ඛ
						ග	[g]	ග
						ඝ	[gh]	ඝ
						ඞ	[ɟ]	ඞ

Figure 7. Genera of the family Brahmic.

of this family are named in Figure 1, and there are other alphabets used in southern and southeastern Asia which will probably form other genera belonging to it. Since it is not convenient to match up their corresponding letters, the alphabets which exemplify the family Brahmic in Figure 7 are simply shown alongside one another, each in its own alphabetical order.

[tʃ] [tʃh] [dʒ] [dʒh] [r] [r̥] [tʰ] [d̥] [d̥h] [n̥] [θ] [θh] [ð] [ðh] [ŋ] [p̥] [p̥h] [b̥] [b̥h] [m̥] [l̥] [l̥h] [v̥] [s̥] [s̥h] [h̥] [ʃ]

च च्ह ज ज्ह र र्ह त त्ह द द्ह न थ थ्ह ध ध्ह ङ ल ल्ह व व्ह श श्ह ह ह्ह

[tʃh] [dʒ, dz] [dʒh] [r̥] [tʃh] [d̥] [d̥h] [n̥] [r̥] [tʃh] [d̥] [d̥h] [ŋ̥] [p̥] [p̥h] [b̥] [b̥h] [m̥] [l̥] [l̥h] [v̥] [s̥] [s̥h] [h̥] [ʃh]

च़ च़्ह ज़ ज़्ह ऱ ऱ्ह त़ त़्ह द़ द़्ह ऩ थ़ थ़्ह ध़ ध़्ह ङ़ ल़ ल़्ह व़ व़्ह श़ श़्ह ह़ ह़्ह

क क़ न. क. ङ ङ़

[tʃ] [tʃh] [dʒ] [dʒh] [r] [r̥] [tʃh] [d̥] [d̥h] [n̥] [θ] [θh] [ð] [ðh] [ŋ] [l̥] [l̥h] [v̥] [s̥] [h̥]

क़ क़्ह न. क़. ङ़ ङ़्ह

Some Further Definitions

Multiple alphabets for a language ; varieties of an alphabet. It is no uncommon thing for a language to be written by different alphabets at different times in its history, and it occasionally happens that a language is written by more than one alphabet at the same time. If these alphabets belong to the same script and there are only slight differences between them, they may be described as varieties of the same alphabet. For example, the Russian orthographic reform of 1918 dropped from that alphabet four letters, two of which were rarely used. The present alphabet, which was established by that

reform, and the alphabet used before 1918 should be regarded as two varieties of the same alphabet, and may be referred to as the Russian (1918 variety) alphabet and the Russian (Peter I variety) alphabet.

If, however, the alphabets of a language belong to different scripts, they must be regarded as different alphabets, and may be distinguished by annexing to their names the names of taxa to which they belong. The two alphabets which have been used to print the German language during this century belong to different scripts of the same genus; they may therefore be distinguished by the names of their scripts as the German (Neoroman) alphabet and the German (Fraktur) alphabet. The two alphabets which are presently used to write the Serbo-Croatian language belong to different genera of the same family. Because of the unusual situation in Serbo-Croatian, they are usually called the Serbian alphabet and the Croatian alphabet, but they may be distinguished more systematically as the Serbo-Croatian (Cyrilloid) and Serbo-Croatian (Romanoid) alphabets, using the names of their respective genera. The alphabets by which the Turkish language is now written and was written before its reform in 1928 have little in common except the need to correspond to the phonemes of Turkish. They may be distinguished as the Turkish (Romanoid) alphabet and the Turkish (Araboid) alphabet.

Sum of a script. This is the set of all letters which occur in at least one of the alphabets of the script. From a printer's point of view, the sum of a certain script is the set of letters for which he must have types (or matrices) in order to be prepared to print in any and every language which has an alphabet of that script.

Quasi-alphabet. There are some systems of written marks which seem to be alphabets—in that they have graphemes, an ordered arrangement, correspondences between their graphemes and certain kinds of speech sounds, and basic shapes for the characters which embody their graphemes—but which are not alphabets because they do not belong to any language. These systems of characters may be called quasi-alphabets. They include, among others, systems of transliteration, systems of phonetic notation like the International Phonetic Alphabet, and alphabetical proposals such as that made for African languages by the International African Institute (1962). Each quasi-alphabet belongs to its appropriate script, and its graphemes are included in the sum of that script.

Generic ligatures ; specific ligatures. A ligature is treated by a language in one of two ways, depending on how it embodies the letters of that language. If a character is an embodiment of two or more letters of an alphabet, it is a *specific ligature*. If it is an embodiment of only one letter but can be recognized, by people who are well acquainted with other alphabets of the same script, as a compound of two or more other characters, it is a *generic ligature*. In English, ffi (f + f + i) is a specific ligature, while w (v + v) is a generic ligature. A ligature may have different status in different languages. Æ is a generic ligature and embodies a single letter in the Danish language, but in the modern English language it is a specific ligature embodying the letters a and e.

Suggestions for Future Research

The taxonomy presented here is only a preliminary outline which will require much work by many hands to become a useful taxonomy of writing systems.

The three taxonomic levels—scripts, genera, and families—may prove to be too many or too few. Consensus will have to develop on the number of taxonomic levels which can be usefully defined and on the kinds and degrees of similarity which are appropriate at each level.

The sample taxonomy does not accommodate all of the world's alphabets. As other alphabets are added to the taxonomy, new taxa will have to be defined to accommodate some of them. The possibility of including logographic writing systems within this taxonomy should be investigated. Some of the taxa described may have to be moved from one level to another; some may prove to be ill-advised and will have to be broken up, with their members reassigned to other taxa. Some lower taxa which appear here as taxonomic isolates can perhaps be included with others in higher taxa.

A taxonomy of this kind is not the only possible classification which can be made of written languages. If other classifications are made that appear to cut across these taxa, the relationships between those classifications and this taxonomy should be investigated.

Although this taxonomy has been constructed by synchronic comparison of writing systems without regard for their historical relationships, the comparison of this taxonomy with the historical relationships of writing systems may lead to some interesting results.

APPENDIX

The taxa which are shown in Figures 3, 6, and 7 use the same system for positioning their vowel and consonant letters; this method is illustrated in Figure 8 by examples from the Devanagari alphabet of Sanskrit. In all of these languages, written words are divided into graphic syllables; each graphic syllable consists of a single vowel letter, a single consonant letter, a sequence of consonant letters, a single consonant letter followed by one vowel letter, or a sequence of consonant letters followed by one vowel letter. (The letters of graphic syllables are not necessarily written in the same order as their corresponding phonemes are spoken; "sequence" and

Figure 8. Arrangement of vowel and consonant letters in the Devanagari alphabet of Sanskrit.

इव प्रति देवास्याम्
[i va pra ti de vā sjā m]

Example of the written language, with its pronunciation divided into syllables.

इव प्रति देवास्याम्
[v pr t d v sj m]

Characters corresponding to single consonants or groups of consonants.

इव प्रति देवास्याम्
[i a a i e ā ā] NO VOWEL

Characters corresponding to vowels.*

प्र स्य
[pr] [sj]

Ligatures: characters corresponding to groups of consonants.

प्र प्र स्य स्य
[p r] [s j]

Decomposition of ligatures.

*The vowel [a] is represented by the absence of any vowel letter; the absence of a vowel phoneme in a syllable is represented by the mark shown here.

similar terms in this description refer not to the letter graphemes but to their corresponding phonemes.)

The marks which write the letters of these languages are of two sizes: large and small. Every letter has its own large mark; some letters also have small marks, which may or may not be partially similar to their large ones. Whenever a graphic syllable consists of only one letter, whether a vowel or a consonant, that letter is written by its large mark. But whenever a graphic syllable consists of a consonant letter followed by a vowel letter, the consonant of that syllable is written by its large mark and the vowel is written by a small mark located in a definite place within the space around the consonant mark—the vowel mark's shape and its location with respect to the consonant mark being equally essential.¹⁶ Whenever two or more consonant letters occur in the same graphic syllable, they may be written in one of three ways, depending on the letters and the language: the consonants may be written as a ligature, the parts of which may or may not be similar to the letters' own large or small marks; one consonant (typically the first) may be written by its large mark, with the other consonants being written by small marks placed somewhere around that large mark; the large marks for all of the consonants may be written in the same order as their corresponding phonemes are spoken. Whenever a graphic syllable consists of two or more consonants and a following vowel, the consonant letters are written as they would be if there were no vowel in the syllable, and the vowel letter is written by its small mark positioned around the whole consonantal assemblage as it would be positioned around a single consonant's large mark.

For many of these languages, there is a vowel letter which has no small mark and which, when it occurs in a syllable with a consonant letter, is written by the absence of any small vowel mark. In such a language, a syllable that contains no vowel may be written with a special mark located like a small vowel mark.

In Figures 3, 6, and 7, the small marks used to write vowels that are in the same syllables as consonants are listed in the columns headed "other basic shapes." To show the required locations of these small vowel marks, the position of the mark embodying the syllable's consonant is represented in the illustration by a gray square, and the vowel mark is shown in its proper place with respect to that square. Figures 3, 6, and 7 do not show the ligatures or the small consonant marks which may be used when more than one consonant letter occurs in one syllable.

16. This method of representing vowel phonemes is called Type III vocalization by Gelb (1963, pp. 197–98).

REFERENCES

- Arden, A. H. *A progressive grammar of the Telugu language*. (4th ed.) Madras: Christian Literature Society, 1955.
- De Laubenfels, M. W. Trivial names. *Systematic zoology*, 1952, 2, 42-45.
- Diringer, D. *The alphabet*. (3rd ed.) New York: Funk & Wagnalls, 1968. 2 vols.
- Downie, R. A. Languages of the world. *Monotype Recorder*, 1963, 42, no. 4.
- Ferguson, C. A. Diglossa. *Word*, 1959, 15, 325-40.
- Gelb, I. J. *A study of writing*. (Rev. ed.) Chicago: University of Chicago Press, 1963.
- Giljarevskij, R. S., and Grivnin, V. S. *Opredelitel' Jazykov Mira po Pis'mennostjam*. Moskva: Akademija Nauk SSSR, 1964.
- Gill, E. Typography. In P. A. Bennett (ed.), *Books and printing: a treasury for typophiles*. Cleveland: World Publishing Co., 1951.
- Goudy, F. W. *Typologia*. Berkeley: University of California Press, 1940.
- . *The alphabet and elements of lettering*. Berkeley: University of California Press, 1942.
- Haas, Mary R. *The Thai system of writing*. American Council of Learned Societies, Program in Oriental Languages, Publication Series B—Aids—No. 5. Washington, D.C.: A.C.L.S., 1956.
- Hanna, P. R., et al. *Phoneme-grapheme correspondences as cues to spelling improvement*. U.S. Office of Education publ. no. OE-32008. Washington: Government Printing Office, 1966.
- Herrick, E. M. *A linguistic description of roman alphabets*. Hartford Studies in Linguistics 19. Hartford, Conn.: Hartford Seminary Foundation, 1966.
- . The linguistic structure of written language. Forthcoming in A. Makkai (ed.), *Linguistics at the crossroads: 1972*. The Hague: Mouton & Co.
- Hoch, E. Commentary: linguistics, "writing", and typography. *J. Typog. Res. (now Visible Language)*, 1968, 2, 379-382.
- Hockett, C. F. *A course in modern linguistics*. New York: Macmillan, 1958.
- International African Institute, *Practical orthography of African languages*. Memorandum I. Oxford: Oxford University Press, 1962.
- Jensen, H. *Sign, symbol and script*. (3rd ed.) Translation by George Unwin. New York: G. P. Putnam's Sons, 1969.
- Kaldor, I. L. The genesis of Russian *grazhdanskii shrift* or civil type. *J. Typog. Res. (now Visible Language)*, 1969, 3, 315-344; 1970, 4, 111-138.
- Mill, J. S. *A system of logic, ratiocinative and inductive*. (8th ed.) New York: Harper & Brothers, 1881.
- Mountford, J. "Writing" and "alphabet". *J. Typog. Res. (now Visible Language)*, 1968, 2, 221-232.
- Nayak, H. M. *Kannada: literary and colloquial*. Mysore: Rao and Raghavan, 1967.
- Noordzij, G. Broken scripts and the classification of typefaces. *J. Typog. Res. (now Visible Language)*, 1970, 4, 213-240.
- Tanselle, G. T. The identification of type faces in bibliographical description. *J. Typog. Res. (now Visible Language)*, 1967, 1, 427-448.
- Trager, G. L. *Language and languages*. San Francisco: Chandler Publ., 1972.