Translating non-Latin scripts into type

Fiona Ross

Corrections

p. 83, 2nd text ¶, 1st sentence, should
   read: 'The reproduction of a com-
   plete non-Latin text dates back to
   AD 868 with the Chinese translation
   of the Diamond Sutra printed by
   xylography.'

Fiona Ross

Translating non-Latin scripts into type

The idea for this article arose from an illustrated talk of the same title given at ATypI 1997 held at The University of Reading, UK.

Non-Latin scripts cover a huge range of writing systems; this paper refers only to those scripts that best exemplify the issues under discussion, and those with which the author is most familiar. Fiona Ross draws on her long experience of working on non-Latin typefaces at Linotype UK. The essay is necessarily an overview, as each issue mentioned could merit an entire article.

Until the 1990s the process of typefounding has rarely, if ever, been performed by one person.[1] Now font tools allow a single individual with a personal computer the facility to design a font and compose with it without recourse to professional typefounders. In the field of non-Latin types the results have not been laudable, and this method is still eschewed by the major font manufacturers. Rendering a non-Latin script in type is a multi-faceted task. It raises issues that traverse specific disciplines; issues which are not necessarily linked to a given technology, and some of which are not necessarily peculiar to non-Latin types.

Since the earliest production of non-Latin fonts, the task of defining the character set has been the foremost consideration. This seemingly straightforward task becomes a complex issue when considering scripts which have no finite character set. North Indian scripts number amongst these. Devanagari, for example, is a script used to represent many languages, including Hindi and Marathi. Not only do we find stylistic differences in the preferred letter formation of Devanagari characters when drawing fonts for these languages, but both are living languages which take in loan words as they continue to evolve. These loan words may contain phonemes or combinations of phonemes that are not inherent in the Indian phonological system, and thus new characters are required to represent foreign sounds. Words in the past decade that have introduced new letter combinations have included transliterations for *street*, *tuttifrutti*, and the like.

The issue of determining the font synopsis is further clouded by the fact that most non-Latin fonts exceed 256 characters. As intimated above in the case of Indian scripts, this excess was usually occasioned by letterforms comprising combinations of basic phonemes, be they clusters of consonants, or consonants conjoined with vowels to produce ligatures.[2] This characterstic only impeded foundry types in as much as the costs of fount manufacture and composition became prohibitive, which in turn inspired the notion of dissecting characters into separate portions and identifying common components which could be used in a multiplicity of situations. This practice culminated in the unreadable distortions visible in the Linotype hot-metal non-Latin fonts where 96 sorts might be the required physical limitation for speedy newspaper composition of, say, the Bengali script, which was usually composed with 510 foundry sorts.

Today's technology seems to operate in units of 256 sorts and, barring such obvious exceptions as Chinese, a passable font may be constructed from this number. Some scripts like Gurmukhi require

1. Despite assertions by such people as Charles Wilkins (1749–1836) to the contrary. See Ross, 1999, 10–12.
2. Or, in the case of Arabic, contextual letterforms and ligatures; see figure 1. See also Saunders, 1997, 30–1.

TÜRKISCH=ARABISCHER KASTEN

Figure 1. Case lay of Arabic–Turkish, D. Stempel AG,
Arabisch Turkish und ihnen verwandte Sprachen,
(Frankfurt, 1922)

Figure 2. Telugu ligature table composed with 310 sorts.
[typeface: Tamara]

क + क = क्क	क्ष + म = क्ष्म	ग + ध + व = ग्ध्व
क + ख = क्ख	क्ष + य = क्ष्य	ग + न = ग्न
क + च = क्च	क्ष + व = क्ष्व	ग + न + य = ग्न्य
क + ज = क्ज	क + स = क्स	ग + ब = ग्ब
क + ट = क्ट	क + स + ट = क्स्ट	ग + भ = ग्भ
क + ण = क्ण	क + स + ड = क्स्ड	ग + भ + य = ग्भ्य
क + त = क्त	क + स + त = क्स्त	ग + म = ग्म
क + त + य = क्त्य	क + स + प + र = क्स्प्र	ग + य = ग्य
क + त + र = क्त्र	क + स + प +ल = क्स्प्ल	ग + र = ग्र
क + त + व = क्त्व	ख + ख = ख्ख	ग + र + य = ग्र्य
क + थ = क्थ	ख + त = ख्त	ग + ल = ग्ल
क + द = क्द	ख + न = ख्न	ग + व = ग्व
क + न = क्न	ख + म = ख्म	ग + स = ग्स
क + प = क्प	ख + य = ख्य	घ + न = घ्न
क + प + र = क्प्र	ख + र = ख्र	घ + म = घ्म
क + फ = क्फ	ख + व = ख्व	घ + य = घ्य
क + म = क्म	ख + श = ख्श	घ + र = घ्र
क + य = क्य	ग + ग = ग्ग	ङ + क = ङ्क
क + र = क्र	ग + घ = ग्घ	ङ + क + र = ङ्क्र
क + ल = क्ल	ग + ज = ग्ज	ङ + ख = ङ्ख
क + व = क्व	ग + ण = ग्ण	ङ + ग = ङ्ग
क + व + य = क्व्य	ग + द = ग्द	ङ + घ = ङ्घ
क + श =	ग + ध = ग्ध	ङ + म = ङ्म
क + ष = क्ष	ग + ध + य = ग्ध्य	ङ + य = ङ्य

Figure 3. Page 1 (out of 6) of Devanagari table of conjuncts
(consonantal clusters). [typeface: Linotype Devanagari]

even fewer sorts. But to ensure good typography, the facility of merging font units to allow for much larger character sets (in a way that is transparent to the user) is a feasible and a necessary solution. It is yet to become universally available. Therefore the make-up of the font repertoire, including numerals, punctuation, and typographic symbols as well as the letterforms, requires careful analysis. The unfortunate, but not uncommon, practice of replicating font synopses of the past, which were constrained by previous technologies, is often inappropriate to current typographic possibilities. Past synopses are merely useful informants.

Thus the first task in defining a character set is the listing of all the letterforms that need to be represented. The next consideration is to determine how these letterforms are to be produced. This is inevitably connected with the composing technique: for instance, should the repertoire include vowel signs attached to host characters, or can software position these accurately? Is space saved by using common components (half-forms), or will the resulting combinations appear contorted?

A crucial fact to recognize in non-Latin composition is that the character repertoire need not equal the number of characters that can be output. A font of 256 glyphs could generate 500 characters or more, depending on the original drawings (figure 2). And as indicated earlier, a method, or at least a policy, should be devised for dealing with the almost inevitable demand for new sorts, for instance by including half-forms which may perform the task adequately in certain typestyles.

To ensure the best possible typographic representation of a writing system, this stage of compiling the character set can necessitate much research. But a flexible approach is also required: as the design of a typeface progresses, new problems may occur or new solutions appear with regard to particular letterforms. Depending on the brief, and any constraints or new developments, the final character set for a given font may not be device-dependent, nor software-dependent, but it could be typestyle-dependent. Script reform committees have at times attempted to standardize character sets, but even when their findings have received government ratification their recommendations have rarely been implemented.

During the history of Devanagari typefounding, a font produced by the Serampore Mission Press in 1836 had 1000 sorts; in 1967 the fonts of the renowned Nirnaya Sagar Press and Jawaj Dadaji typefoundry comprised 608 sorts; the hot-metal Linotype fonts, which were justifiably criticized, were limited to 90 principal sorts, 34 auxiliary characters, and a few *pi* sorts; the subsequent Linoterm film fonts contained 110 sorts each. The first digital Devanagari font,[3] which was produced in the late 1970s for the Linotron 202, contained approximately 300 sorts with almost yearly additions of a handful of characters (figure 3). The first successful Bengali font of 1778 contained 170 sorts; Bengali typography eventually reached a standard of 510 sorts; the first digital Linotype Bengali fonts had some 300 sorts. Other font histories have their own tales to tell.

Another issue of concern to a font developer is whether the script is joining or non-joining. The prevalence of non-Latin fonts that join is clearly dictated by the nature of the scripts. But this characteristic is

3. Linotype Devanagari Light. The Monotype Corporation digitized its Devanagari fonts in 1980 for use on the Lasercomp, (Shaw, 1980, 28).

probably no more problematic than for the designer of a joining Latin font. However, some of the earliest non-Latin digital fonts were offset to permit large amounts of kern, which caused a problem in some scripts when a character with a longer joining stroke on one side, due to the offset, was required to be set in isolation. Problems can also occur when joining strokes are curved; thus particular attention needs to be paid to the overlap and the 'hinting' of the character outline for low-resolution output. (One Punjabi newspaper proprietor, on acquiring digital Gurmukhi fonts, insisted on having the option – and thus the software facility – to retain the gap in the headline formerly visible in metal until his readership became accustomed to the continuous headline.)

A non-Latin font developer may need to contemplate different reading directions, but this issue is more complex in terms of composition than design, particularly in mixed-language setting. It relies on software solutions. It is the character body-height and the multi-level nature of some scripts that exercise type designers more. In order to achieve a good balance of ascenders and descenders, as is desirable in Arabic, or to cater for multi-level characters that may have great depth as well as height, the customary Latin baseline might be jettisoned when drawing and digitizing the characters (figure 4). The composing software may thus have to reconcile different baselines when mixing fonts or, in the extreme case of the Nasta'liq script, handle multiple base-jumps within one word (figure 5). Here again, the type designer and the programmer need to work in tandem.

From the user's point of view, the three- to four-tier nature of the North Indian scripts generally precludes the possiblity of setting solid, i.e. without extra leading (figure 6). If compact setting is required the interlinear spacing needs to be adjudicated by the user, who may opt for alternative character forms (e.g. horizontal rather than vertical joins). The compositor (and the designer) will also wish to know the policy of typeface compatibility held by the non-Latin font vendor.
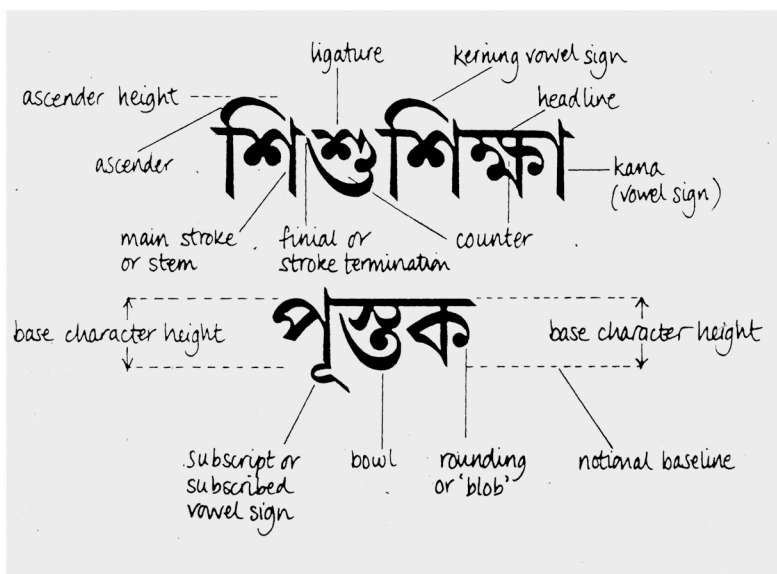

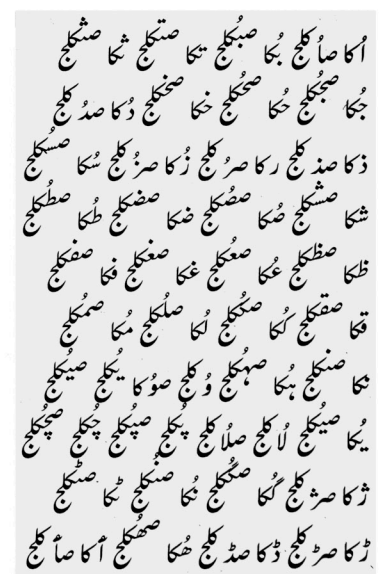Figure 4. Typeface nomenclature with baseline. [typeface: Linotype Bengali]

Figure 5. Prelimary tests for base-jumps in Nasta'liq. [typeface: Qalmi]

Figure 6. Three- to four-tier writing system. [typeface: Linotype Bengali]

BBox: min = 11, -228   max = 984, 690
SideBearings: L = 11   R = 16   Width = 1000
Parts: mt = 23   dt = 52   ct = 222   Total = 297
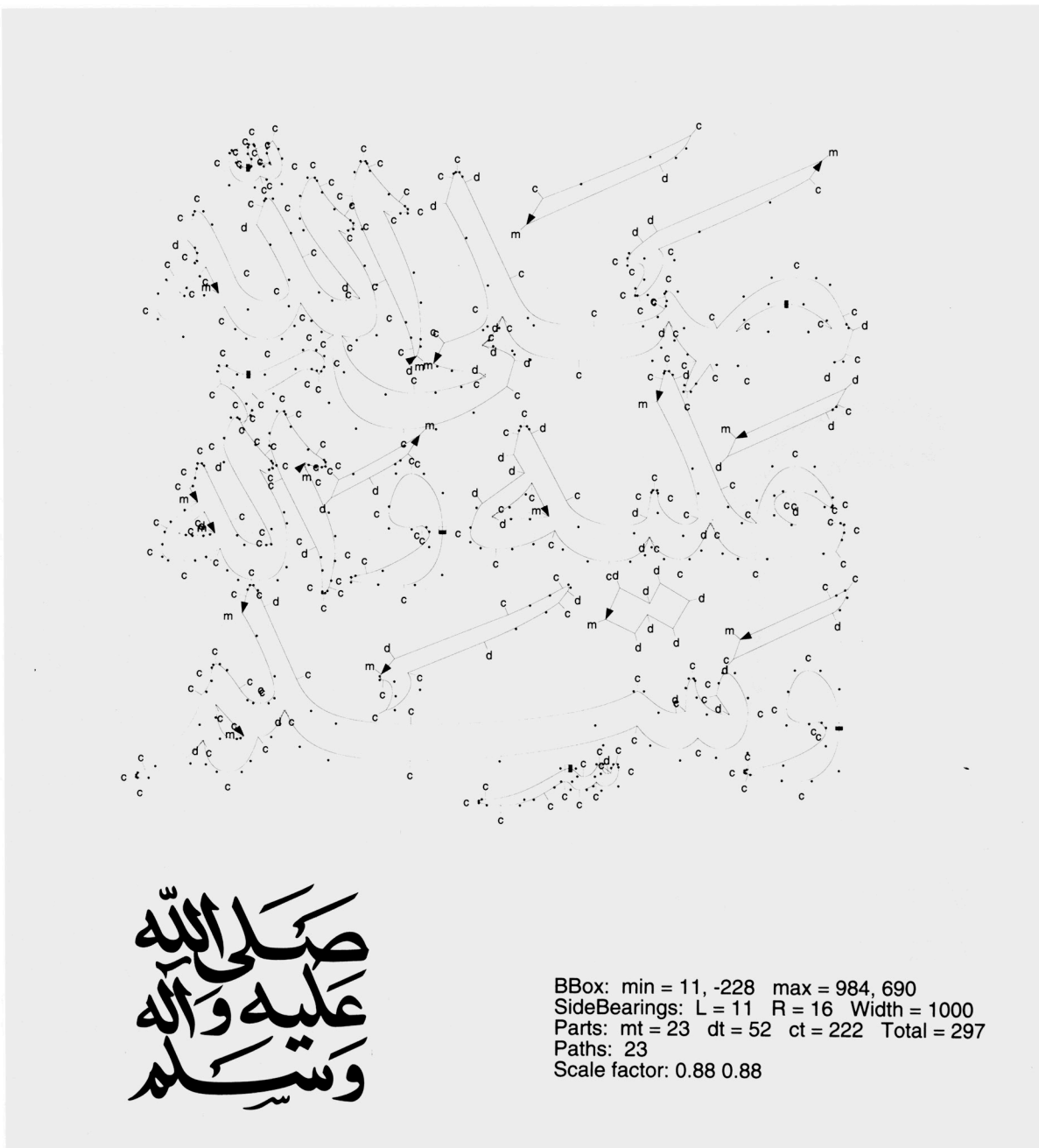Paths: 23
Scale factor: 0.88 0.88

Figure 7. Digital output showing point placements of Farsi version of 'Praise be to Allah'. [typeface: Qadi]

Non-Latin scripts often display a smaller appearing height than Latin. Some font manufacturers recommend that in mixed-script setting the non-Latin font is set two point sizes larger than Latin types to counter this effect. It is not uncommon to see bilingual publications that have failed to recognize this, with the consequent result that some readers criticize the publications for giving prominence (thereby connoting pre-eminence) to the language in Latin type.

Even amongst fonts of the same script, discrepancies in type-size are prevalent. This may be due to a font manufacturer's policy that a designer should not be constrained by the earlier standards of pre-existing designs, but that priority should be given to designing the best possible typeface within the given technology. In the case of non-Latin fonts, such a policy has usually been appreciated by the user, as well as welcomed by the designer.

However, questions regarding compatibility in alignment often elude consideration and can occasion criticism. Generally, the composing software attempts to align fonts along a 'notional baseline' even though it might be preferable for typestyles bearing a 'headline', such as the North Indian fonts, to align along the headline. In practice, font changes never occur mid-word and thus such inconsistencies do not arrest the reader's eye; but when characters occur as superiors, the failure to align along the headline in such scripts is unsatisfactory. An option to do so would be preferable.

Some issues that occupied non-Latin font developers in the past have now been eliminated. Many non-Latin writing systems employ letterforms of far greater complexity than those found in Latin fonts. At times this has caused characters to be composed from several sorts, or in the case of digital fonts to be scanned in more than one pass. Now such letterforms can be designed and digitized as integral characters, stretching the abilities of font editors in the 'hinting' process to render such characters as sharp as possible in all sizes and resolutions (figure 7). Formerly, exceptionally wide characters, such as occur in Tamil and Malayalam, necessitated a reduction in body-height to accommodate them. The issue now is whether the type height should be increased for improved clarity and greater compatibility with other scripts when translating popular designs from previous technologies into new font formats.[4]

The ability to kern elements forms an essential prerequisite for the production of readable non-Latin fonts. Although the Monotype machine could compose kerned characters, it was the inability of linecasting machines to do so that produced the poor typography which characterized non-Latin typesetting in the hot-metal era. The Bengali script, which has large vowel signs that kern either to the left or right, suffered much distortion in hot-metal composition, and the appearance of condensed vowels in daily newspapers created a readership that was subsequently unable to recognize orthodox letterforms (figures 8 & 9).[5]
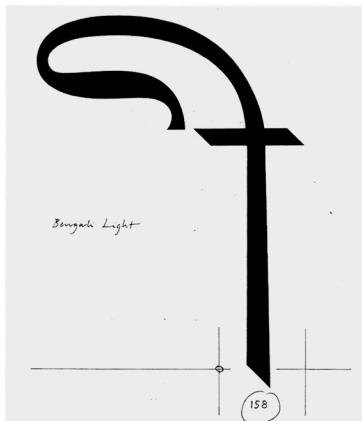

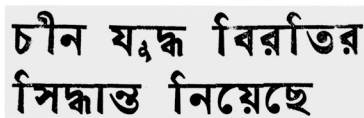Figure 8. Kerning Bengali vowel sign. [typeface: Linotype Bengali]

Figure 9. Text from *Ananda Bazar Patrika* composed with hot-metal Linotype fonts.

4. Character widths still cause problems when designing screen fonts, but this issue is beyond the scope of this essay.

5. The Monotype machine, with its kerning capabilities and its larger character sets, surpassed the Linotype machine in this respect, although character shapes were compromised by width restrictions. Its infrequent use in newspaper composition, however, meant that Monotype fonts exerted relatively less influence on Indian typography, particularly with regard to North Indian scripts.

6. As was practised in earlier technologies, e.g. Monotype hot-metal.
7. *The Linotype Record*, July 1923.

Another feature of the majority of non-Latin scripts is the necessity to position subscripts and superscripts accurately. A font may have 16 'accents' to position above or below 300 characters. The positioning is particularly vital as poor placement impairs readability, whilst incorrect positioning can alter the meaning of a word. In the digital era good placement is achieved by software using x and y co-ordinates to locate the correct placement in relation to each character (figure 10). For some font vendors this causes problems in that the fonts will only function with specific software. An alternative solution is to produce multiple versions of the same accents in various positions which can be used with different characters.[6] The problems here are that the results are not always accurate; that this method extends the already large character set; and that some sort of software is still required to select the appropriate version. New font formats may be able to embed this sort of information in the font.

However, non-Latin scripts demand other software requirements of varying degrees of complexity. Selecting appropriate contextual forms for Arabic is no longer regarded as onerous, but some scripts, such as Telugu, have particular spacing requirements (figure 11); others require base-jump selection; others, like Thai, incur complex word-breaking logic. The designer cannot prepare finished artwork until there is a clear indication of how the font will be composed and with what software utilities. Answers are needed to such questions as: Are ligatures to be designed as integral characters? Are the vowel signs positioned to work from the left or the right?

Keyboards and their layouts have always played a vital part in non-Latin typesetting. There are records of Arabic keyboards being manufactured since 1911 for newspaper production, where keying speeds are so crucial.[7] Now with digital technology the keyboard layout is much less entwined with the font layout; and software logic has

Figure 10. Part of an Arabic diacritical placement test set at low resolution. [typeface: Yakout]
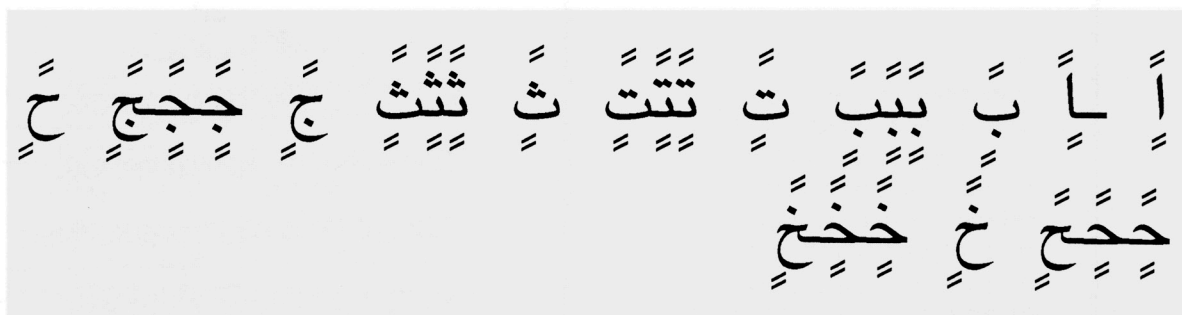

Figure 11. Short Telugu text to illustrate spacing. [typeface: Tamara]

8. Based on the Indian phonological writing system. See Ross, 1999, 216–220.

9. *Specimen des caractères, vignettes, armes, trophées et fleurons de l'Imprimerie Royale* (Paris, 1819).

rendered redundant the huge keyboards of the past which required specific hardware. Keyboards for composing Indian scripts have been much simplified since Linotype-Paul Limited invented the phonetic keyboard in 1978.[8] But the type designer still needs to be aware of how characters are to be accessed and formed when preparing artwork.

One issue that appears to receive insufficient attention in the case of non-Latin fonts is the design of the lettershapes themselves. Font developers have tended to focus on the apparent technical problems rather than on readability. When translating non-Latin scripts into type, the objective should be the same as that of a translator of languages: to form an imperceptible medium of communication between the writer and the reader. A non-Latin font should represent the original language in a form that is immediately comprehensible, as well as pleasing, to the recipient.

In designing a non-Latin typeface, a script's cultural heritage, as is visible in monumental work, engravings and the like, provides vital clues as to the suitability, and ultimately readability, of a design. A good example of the failure to take note of cultural traits is afforded by the specimen of Bengali types printed by the Imprimerie Royale in Paris in 1819 (figure 12).[9] At first sight it appears to be a well-presented type specimen. However, this font of uncertain provenance does not bear close scrutiny. The text is decipherable but not readable: the proportions of the characters are incorrect; the weight distribution is disparate; the relationship of the diagonal strokes to the vertical strokes does not accord with Bengali penmanship; and the spacing is uneven. Each letterform could be criticized, but suffice it to say that it is a poor representation of the Bengali script. The Imprimerie Royale was far

Figure 12. (below left) Bengali types of the Imprimerie Royale. From *Specimen des caractères, vignettes, armes, trophées et fleurons de l'Imprimerie Royale* (Paris, 1819). St Bride Printing Library.

Figure 13. (below right) Eighteenth-century cursive Bengali manuscript (cursive hand). From BL MS Add 5660E – An obligation'. British Library.
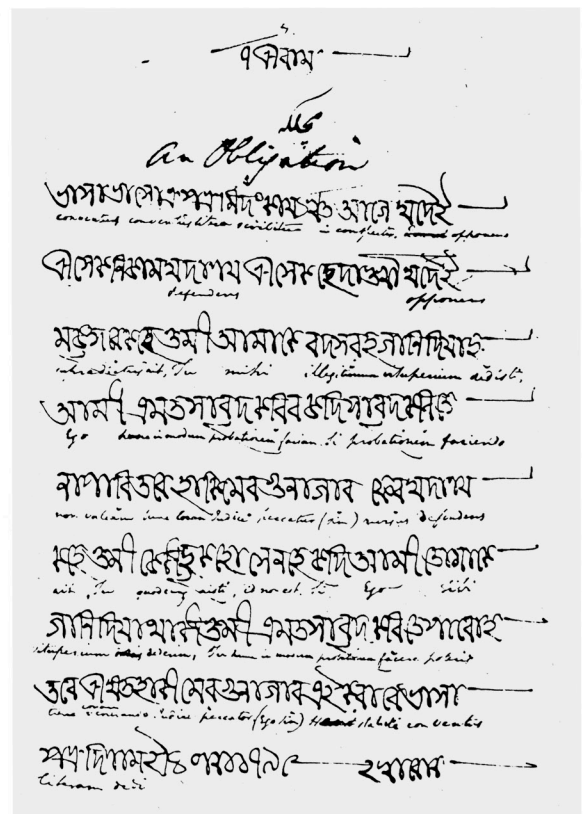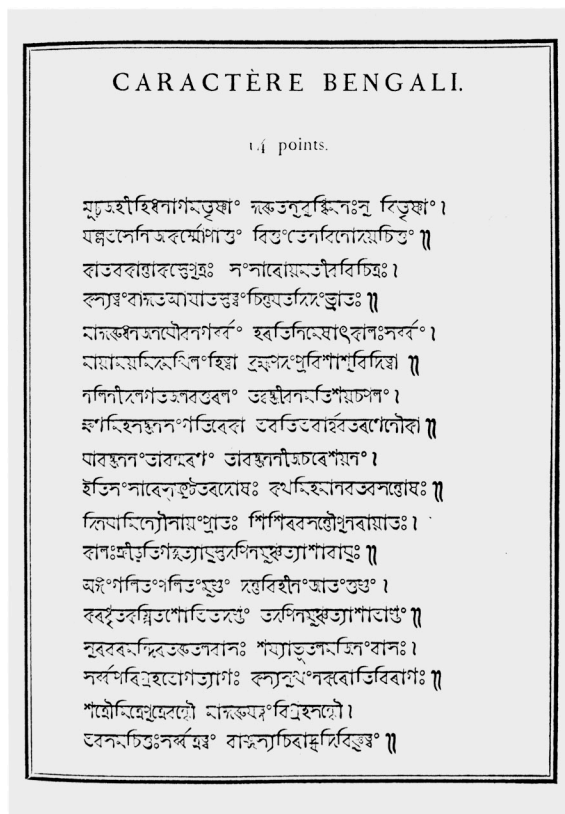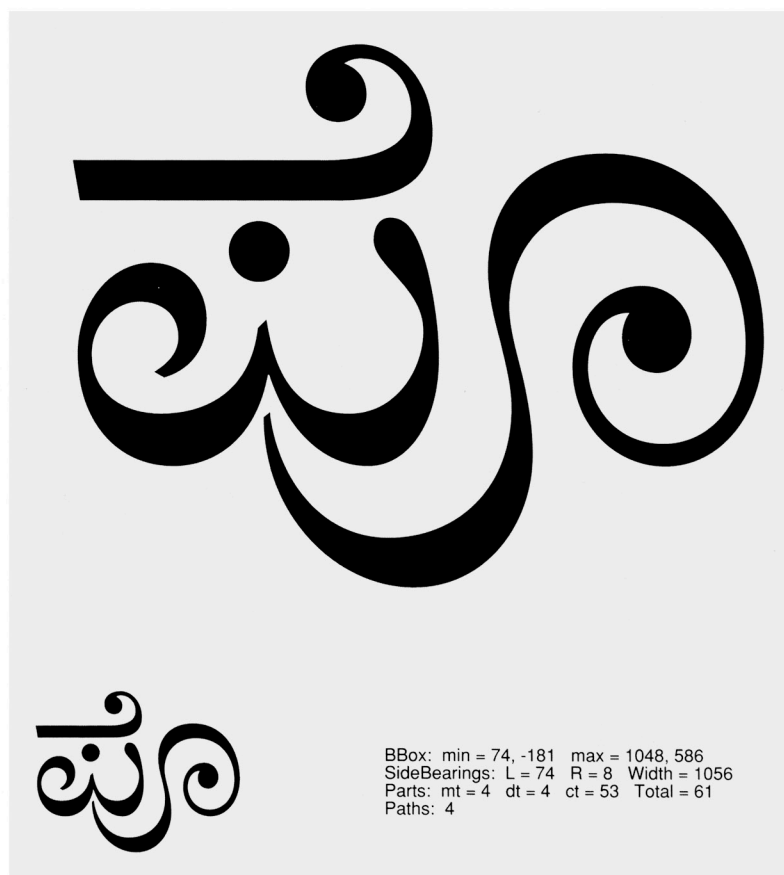
removed from Bengal and the typecutter probably had no knowledge of Bengali language or culture.

It is useful to consider the handwritten script which the Imprimerie Royale was attempting to render into type; to note that there existed two distinct calligraphic styles: a cursive hand (figure 13) and a decorative hand. But such hands do not appear to have informed the French font. When reproducing a script in type the designer needs to be aware of the calligraphic heritage, to review current handwriting and lettering practices, and to reflect that such a script represents a living language. The histories of Arabic and Japanese type design show a continual rapport between calligraphers and type designers that is absent in some other scripts.

The reproduction of a complete non-Latin text by means of moveable type dates back to AD 868 with the Chinese translation of the *Diamond Sutra* printed by xylography.[10] The manufacture in fifteenth-century Korea of fonts comprising 100,000 sorts confirms that the requirement for printing non-Latin scripts has always existed alongside that for Latin-based languages. Many non-Western scripts boast a rich typographic tradition, notwithstanding the paucity of such fonts in major type libraries. For more than a century, foundry types of note have emerged from indigenous foundries of different continents; and foundry types have continued in use as alternatives to their inferior counterparts in hot-metal and early photocomposition.

The typographic heritage of a given script can inform the designer as to the acceptability of certain styles. The informed viewer of typeface specimens and other published material can distinguish between

Figure 14. Kannada ligature. The rounding of the finials is vital to identifying the character. [typeface: Kesari]


```
BBox: min = 74, -181   max = 1048, 586
SideBearings: L = 74   R = 8   Width = 1056
Parts: mt = 4   dt = 4   ct = 53   Total = 61
Paths: 4
```

fonts of quality for bookwork and legible, robust newspaper faces; between classical fonts suited for religious texts and those fonts that now appear antiquated, or simply dated. The type developer can take stock of a script's typographic history but new type designs need not be modelled on typographic antecedents, particularly as former type-founding limitations have been removed. Too many current fonts display the legacy of the pre-digital eras. We now encounter distorted and distended, yet orthographically correct, characters that have been generated by fusing different, often disharmonious, components. Their use produces words that are altered from their readily recognizable shapes, thereby compromising readability. This is a critical issue when it concerns languages used in areas of low literacy.

The linguistic aspects of a script also inform our visual criteria for assessing a design. The ability to distinguish between embellishments and identifying features is vital (figure 14). It is also crucial to gauge the relative proximity of sorts to avoid obscurities in meaning – this is often overlooked in Gujarati where, in an ambiguous context, over-tight spacing can unintentionally imply a phonemic combination. Furthermore, knowledge of a language, such as Sanskrit, assists in the recognition of the stylistic features that act as visual signposts to its morphology.

Nonetheless, the same standards of design need to be applied as for a Latin font, even if different visual criteria are required. The application of weight to embolden a font, the stress of the counters, and the deliberately flattened sections on curves may be script-related, but an experienced designer would recognize, even if unconversant with the writing system, whether good design standards have been met. Non-Latin designs still require that there should be a constancy of similar

Figure 15. A 60 pt Tamil proof marked up for correction. [typeface: Samanti]
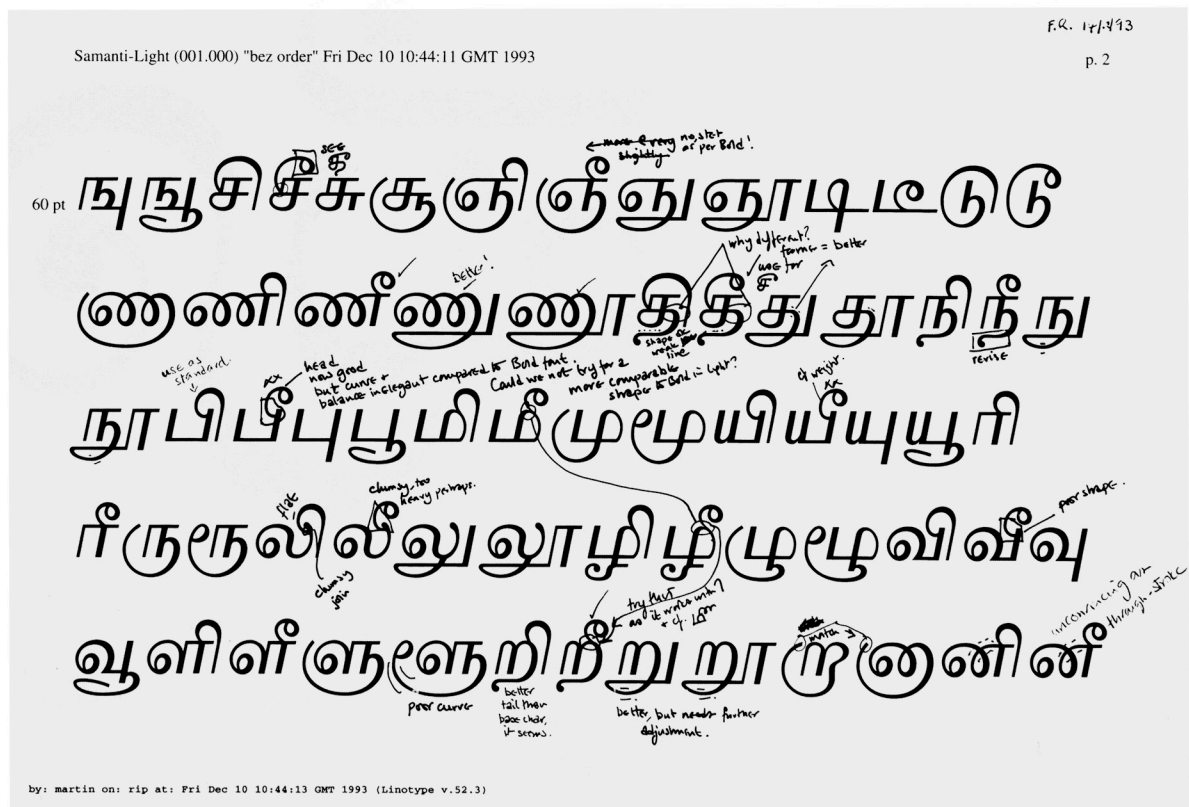
11. Skills which rarely reside in an individual, but which may be best effected either by in-house teamwork or by 'networked' individuals.

12. QuickDraw GX is a product of Apple Computer, Inc; OpenType is being developed by Microsoft Corporation and Adobe Systems Inc.

elements; that the transitions from curves to straight strokes should be smooth; that the weight distribution should be even, and so forth (figure 15).

As with previous typefounding technologies, the successful rendition of a non-Latin script in type depends upon a multiplicity of skills – technical, linguistic, and artistic.[11] The relative lack of innovation in non-Latin type development during the last ten years may be due in part to the diverse, but seemingly discouraging, issues necessarily encountered in the production of non-Latin fonts. Flagrant piracy has been a futher disincentive for font manufacturers to commission what may be perceived as more costly, yet less marketable, designs (figures 16 & 17). However, the increasing demand for vernacular fonts of sufficient quality to further literacy campaigns, to improve regional and multilingual typography, and to supply multimedia projects, combined with the current, more flexible and accessible font production techniques, should ensure the feasibility of commercially viable, high-quality non-Latin fonts. The typographic capabilities offered by new font technologies such as QuickDraw GX and OpenType,[12] which appear so advantageous to vernacular scripts, herald an era of exciting possibilities for non-Latin composition. With the approach of the third millennium there can be little excuse merely to approximate a writing system.

Figure 16. (left) Specimen of text composed with digital Linotype Bengali fonts designed by Tim Holloway and Fiona Ross. Initially criticized for being retrogressive by reintroducing orthodox forms, it is now the most pirated Bengali typeface in India, Bangladesh and Britain.

Figure 17. (right) Nasta'liq type specimen. [typeface: Qalmi]

References

Unpublished sources

Manuscripts and records

    British Library: Oriental Collections, Manuscripts

    N.B. Halhed Collection: [Bengali] Additional 5660E.

Linotype Library

    Typographic R&D files from Linotype Limited, Cheltenham, UK

Published sources

Crystal, D. 1997. *The Cambridge encyclopedia of language.* 2nd edn
    Cambridge: Cambridge University Press

Gaur, A. 1984. *A history of writing.* London: British Library

Imprimerie Royale. 1819. *Specimen des caractères, vignettes, armes,*
    trophées et fleurons de l'Imprimerie Royale. Paris

Naik, B. S. 1965. *Typography of Devanagari.* Bombay: Directorate of
    Languages

Ross, F. 1999. *The printed Bengali character: its evolution.* Richmond:
    Curzon Press

Saunders, D. 1997. 'Two decades of change: 1965–1986', *The Monotype*
    Recorder, new series, 10: 26–35

Shaw, G. 1980. 'Printing in Devanagari', *The Monotype Recorder*,
    new series, 2: 28–33

Stempel. 1922. *Arabisch Turkisch und ihnen verwandte Sprachen.*
    Frankfurt am Main: D. Stempel AG