

***Standardization of West African Scripts:
A Fresh Perspective on Linguistics, Typography, and Literature***

January 21, 2011

Presentation to SIGDD79, Yokohama, Japan

Charles Riley, Yale University Library

Introduction

Among the many misconceptions about language in Africa can be counted a widely held notion that no written language has emerged from the entire continent over the course of the past five thousand years. Nothing could be further from the truth, of course, as even a cursory glance at history will remind us that dynastic Egypt had a hobby of scribbling on its walls, and lay claim to a large corner of Africa for several millenia. But the story of how the relation between writing and African languages has evolved goes much deeper, tracing through Ethiopia and medieval Timbuktu, and intensifies over the past four hundred years, much of which was marred by colonialism and the tragedy of the Atlantic slave trade.

Any presumption of African languages being only spoken, not written, then, can be put to rest, especially in considering those that have been widely used in the last hundred years. Surveys of the total number of African languages have put the figure at somewhere approaching two thousand; of those, perhaps five hundred or so have a functioning practical orthography actively used in creating texts and some degree of publishing. Latin script and Arabic script usages predominate, and for most of southern and eastern Africa the Latin alphabet is used without any major modifications. The main exception to this is in the case of the Khoesan languages, where click sounds are represented using characters from the International Phonetic Alphabet (IPA).

By and large, though, it is western and central Africa that have experienced the most contortions of Arabic- and Latin-based orthographies to fit their phonetic inventories of consonants, vowels, and tones. Adaptations of both scripts, as ‘ajami’ and the IPA respectively, have each been effective for individual languages; typically it is the modified Latin alphabets that are used for publication. While not ubiquitous, technologies to

support input and data management for practical usage of languages using these scripts on the web and in desktop publishing have become more readily available in recent years.

The story might have ended here, but given the multilingual nature of many west African societies, where intermarriage between ethnic groups is quite common and a certain arbitrariness to the way in which national borders were drawn, the linguistic landscape has remained dynamic as a field of innovation, particularly in the past two hundred years. Eric Wolf reminds us in *Europe and the People Without History* that even cultures limited to oral tradition were never static and unchanging, but contributed to and participated in the process of historical change; this was certainly true in the case of the coast of Upper Guinea, where several of the indigenous scripts I am discussing today emerged. Narratives of how they were invented often point to inspirations from dreams (Vai, Garay), but at a deeper level, the act of writing can also be seen as a conscious act of resistance to colonial power, as in the case of Loma and Bamum, where the teaching of the scripts came to be actively suppressed. At a practical level, at least three of these scripts: the Garay script of Assane Faye designed primarily for Wolof, Frédéric Bruly Bouabré's script designed primarily for Bété, and N'ko, invented by Souleymane Kanté, were each intended to serve the needs of more than one language, to facilitate written communication over a region and between ethnic groups. In the cases of N'ko and Garay specifically, these scripts can be seen to function in part as a simplified shorthand for Arabic, thereby allowing sounds from Arabic to be represented adequately within a script that also, through the use of diacritics, provides for sounds from European and African languages, in ways that unmodified Latin and Arabic scripts might not.

The centrality of writing to whether a language is perceived as 'modern', and the nuance that language brings to the understanding of a culture, cannot be understated. Within the academy, attention to the emergence of written forms of language in this region has typically been the domain of individual researchers: David Dalby, Diane Oyler, P. E. H. Hair, Christopher Wyrod, Saki Mafundikwa, Konrad Tuchscherer, Ruth Stone, Sylvia Scribner, Artiom Davydov, Valentin Vydrin, Svend Holsoe, Robert Leopold, and Théophile Obenga, to name just a few. In some cases, studies have been institutionalized at least at the level local to where the languages are used—there is a lecturer in Vai at the University of Liberia; there are classes taught in Bamum with reference to manuscripts housed at the palace in Foumban, Cameroon; and publishing activity in N'ko has continued at a steady pace in Guinea and Mali. With this background in mind, we can proceed to discuss several of these scripts in more detail, noting the progress that has been made so far in establishing support for their use on the web, along with a view toward what needs remain to be covered.

Bamum (A-ka-u-ku)

The Bamum script went through seven distinct phases of development, beginning ca. 1896 with a set of ideographs constructed by Sultan Ibrahim Njoya in close consultation with associates at the palace of Foumban, in Cameroon. The character set began to expand in successive stages, before being reduced to a repertoire of 80 syllabic base characters, two diacritics, ten digits, and six punctuation marks, largely arriving at its current form ('A-ka-u-ku') between about 1910 and 1920.

A key challenge in establishing the Unicode encoding proposal for Bamum was to decide how many of the archaic forms from early and intermediate stages were necessary to be preserved as individually encoded characters. The repertoire for 'Old Bamum' was settled on a case-by-case analysis of charts from Dugast and Jeffreys (1950) that tracked the development of each character through different phases. The proposed encoding repertoire was submitted to Oumarou Nchare, director of the palace museum at Foumban, and met with his approval.

At the implementation level, one key piece remaining is to provide the rendering support needed to correctly place combining diacritics over the modern base characters, through platform-specific rendering engines including Uniscribe, ATSUI, and pango. Future directions for research should include more complete analysis of primary sources, Bamum manuscripts held at the Musée du Quai Branly in Paris, and to review ongoing research in Foumban, in case there are attestations of identifiable additional characters not yet accounted for in the secondary sources. The character repertoires as known to date are reflected in the Old Bamum and Bamum ranges, as of Unicode 6.0.

Bassa Vah

The Bassa Vah script was largely developed by Thomas Narvin (or Darwin) Lewis, while a medical student in Syracuse, N.Y. in the early 20th century. He was the son of a chief from the Bassa ethnic group in Liberia, and had the opportunity to study in the U.S. Folk traditions trace the Vah script farther back, to at least the middle of the 19th century with Jenni Dirah, but evidence for use this early remains scant.

Lewis used equipment at Lyman Bros. in Syracuse and in Dresden, Germany to typeset and print in the script (1907-ca. 1920). On his return to Liberia, he set up a press that was in use until the 1960's (Abba Karnga, personal communication). Students taught in the script included Abba Karnga, Barni Cheevahn, Jacob Logan, and Representative Thomas H. Greeves (Dalby, 1967).

The script accounts for 23 consonants, six vowels, and five tones. During the course of developing its encoding proposal for the Unicode Consortium, two of the main questions that had to be addressed were (a) whether or not it uses casing, and (b) the names of the five tones as rendered in the Bassa language.

For the first question, sources that refer to the use of casing (upper and lower) can be traced to a font developer who worked on the script in the mid-1990's. His intention was to provide glyphs that preserved both the printed appearance and handwritten forms of the script, and thought that one way to do this might be to split the character set into two cases. The idea met initially with the acceptance of the elders he consulted, but casing was never widely implemented in actual practice. In preparing the encoding proposal, consulting both with the font developer and the elder (Karnga) whose handwritten style he intended to preserve, it was agreed that the encoding would not require the separate casing of lower case characters after all, but that these styles could be preserved as glyph differences between different fonts.

On the second question, it is the nomenclature of different sources in English for the tones that led to the need to identify the tone names as described in Bassa. While one source (Dalby) denoted them as <high>, <low>, <mid>, <mid-low>, and <high-low>; another (UBOA) listed them as <high>, <grave>, <mid-low>, <drag> and <double>, respectively, such that <mid-low> was in use for two different tones, depending on the source. With the help of Abba Karnga and Timothy Slager, these were identified more reliably as <neìn wuḍu dyi>, <ḍo wuḍu ɓoúm>, <kpa wuḍu dyíi>, <gbàìn wuḍu dyi>, and <gbèìn wuḍu mú> (Karnga and Slager, personal communication). The regular positioning of combining tonal markings in the middle of the base characters they modify is a feature apparently unique to Bassa Vah; at least this author does not know of other scripts where this is known to happen.

A preliminary encoding proposal was presented at the last WG2 meeting in Busan, South Korea, in October of 2010.

Bété (of Bouabré)

Beginning in the 1950's, an artist by the name of Frédéric Bruly Bouabré began working on developing a set of stylized ideographs for use primarily with the Bété language of Côte d'Ivoire. His artistry was exhibited internationally over the following decades, and stories written in the Bété script were translated and published. The script itself has not reached wide adoption, but he has been a prolific writer and several manuscripts as written in the script he invented remain to be published as originally written. So far, a non-Unicode font has been developed, and is under review by Bouabré and members of his household, who I met with at their house in the Abidjanais suburb of Yopougon-GESCO in the spring

of 2009. Both Latin-transcribed data and sound recordings of a recitation of syllables are available for research and to develop the encoding proposal, which should be happening over the course of the next year or two.

Garay (Wolof, of Assane Faye)

As with Bété, the inventor of the Garay script is still living and keeps a house full of manuscripts on various subjects. Garay is a right-to-left phonetic script with a set of combining diacritics, and was invented at the dawn of Senegalese independence as Assane Faye heard the speeches of Senegal's first president, Léopold Sédar Senghor, exhorting every citizen to each contribute what they can toward the progress of the country. Faye's works include a translation of the Koran, texts on social ethics and political philosophy, a math textbook, and an illustrated book of folktales. Over the past fifty years, he has worked to teach the script to an estimated 200 people, including women's groups. At the time of my visit in the spring of 2009, he was giving a press conference in which he recounted the history and utility of the script in detail. There has been preliminary work on font development, and an encoding proposal will be prepared shortly.

Kpelle

A script for the Kpelle language of Liberia and Guinea was invented by Chief Gbili of Sanoyea of Liberia ca. 1930. Peter Giddings assisted in the introduction of digits. Historically documented users of the script included Lee-Polu-Mala-Yale, of Zongkai, Chief Bono-Boi of Yanekwele, and one of Gbili's wives, Neni-Tele (Stone, 1990 and Dalby, 1967).

Versions of the script are printed in Lassort (1951), Dalby (1967), and Stone (1990). Dalby rearranges Lassort's data into a chart based on a phonetician's sort order, rather than a sort based on Latin script equivalents. Stone's data includes syllable lengtheners and other characters not seen in the other published material. Unlike other syllabaries, almost every Kpelle character can be used for either of two different related syllables, what Dalby terms 'mutational pairs'. An additional feature is that nearly every character can be represented by either of two visually distinct glyphs. Either glyph in a pair can be used to represent either syllable in the same pair, interchangeably.

Konoshenko (2008) identifies three dialects of Kpelle: Gbali, Standard Guinean Kpelle (SGK), and Standard Liberian Kpelle (SLK). While Gbali has one tonal class and one pitch level opposition, SGK and SLK both have two tonal classes and at least two pitch level oppositions.

Winkler (1997) notes, following Welmers' 1973 analysis, three distinct levels of tone: high, mid, and low, plus a falling glide; tone can have both lexical and grammatical functions.

Dalby stated that the syllabary in its current form may still be incomplete, although the later data from Stone fills in some of the gaps he mentioned. There may be some open questions with regard to the use of diacritics in the script as concerns vowel length.

Loma

The invention of a syllabic script for Loma (or Toma) is credited to Widi Zobo of Boneketa, Liberia, dating to the late 1930's. Interviews facilitated by Jacques Onivogui with Vieux Kugbe Onivogui, Balla Traore, and others in 2009 helped to confirm that the script had been in use as a means of exchanging letters between Loma soldiers during World War II, that it was taught in informally structured classes near the rubber plantations at Boneketa through the 1940's, and that there is a substantial amount of interest in maintaining and spreading its use among speakers of Loma today, on both sides of the Liberia-Guinea border. One complicating factor in working toward its encoding has been the persistence of an unreliable print source, which was misleading in the character identifications it made. It did not receive wide circulation, but influenced one set of digital character images derived from it. Other sources—notebooks maintained by Balla Traore and Pema Toupou, and a chart printed in Dalby (1967) and reprinted in Vydrin (1987), are considered by the community to be more reliable and are more consistent with each other.

Mende (Kikakui)

A right-to-left script known as Kikakui was developed by Kisimi Kamara for the Mende language, circa 1920, in part influenced by his study under Mohamed Turay whose own work doubtlessly contributed to the success that Kamara was able to realize. With respect to encoding, there are some variables that increase the difficulty of providing for an encoding for Mende: (a) sources vary widely on the exact sort order in its particulars, although there are general similarities; and (b) the existence of a repeatable diacritic over digits that indicates the power of ten—as it can in principle be repeated indefinitely, but the combined set of repeated marks over the digit should remain centered with respect to the base character. A preliminary proposal has been drafted, but it has been quite difficult to find literates in the script willing to provide substantive comment on the draft to date.

N'ko

As with Bamum, one of the key pieces of technical infrastructure still needed for N'ko is to work out the proper specifications for rendering diacritic positioning, as well as implementing ligatures and right-to-left behavior. N'ko has been in use since 1949, primarily for the Mande languages of Guinea and Mali, following the prolific work of its inventor, Souleymane Kanté. There are a small number of newspapers actively published in N'ko script, and a growing number of book titles, many of which are published in Egypt by a Guinean professor, Baba Mamadi Diané. Its introduction into Unicode 5.0 was supported in part through UNESCO's Initiative B@bel.

Vai

Of the indigenous West African scripts, Vai is perhaps the best documented, and is among those that have advanced the farthest in terms of digital encoding support and infrastructure. Invented ca. 1830 by Duala Bukele, a full-length manuscript dating to this period written by Bukele's cousin is housed in Harvard University's Houghton Library. Its encoding was introduced in Unicode 5.1, and a system font (Ebrima.ttf) provides for its character range as well as that of N'ko in Windows 7. Even so, there is a long way to go in bringing it up to full speed, with such projects as locale support and refining an online dictionary weighing in the balance.

Other scripts

The full list of African scripts may never be exhaustively described: the Latin alphabet itself, even as we know it, has had some of its origins traced to ancient rock art in pre-dynastic Egypt deserts; stelae of Meroitic inscriptions remain largely undeciphered in Sudan, and mystical symbols from Timbuktu manuscripts and Morocco will likely remain as enigmatic as originally intended by their authors.

That said, substantial work has been completed to successfully encode not only West African scripts, but others such as Tifinagh, Ethiopic, and Osmanya into Unicode. There are a number of remaining scripts such as Mandombe, Masaba, and scripts infrequently used for Hausa and Fula, where progress toward encoding digital standards may also be needed. Certainly maintaining quality input tools, rendering, and locale support across platforms for scripts in active use is an ongoing area for future research and development at the implementation level.

Bibliography

- Dalby, D. “A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Kpelle and Bassa”. *African Language Studies* 8:1-51, 1967.
- Dugast, I. and M. D. W. Jeffreys. *L'écriture des Bamum : sa naissance, son évolution, sa valeur phonétique, son utilisation*. [Duala ? : s.n., 1950].
- Karmo, V. Personal communication, 2008-2010.
- Karnga, A. Personal communication, 2010.
- Konoshenko, M. “Tonal systems in three dialects of the Kpelle language”. *Mandenkan* 44:21-42.
- Lassort, Lassort, A. “L'écriture guerzée”, in *Première Conférence Internationale des Africanistes de l'Ouest, 1945, Comptes rendus*. Vol. 2. 209-215, 1951.
- Stone, R. “Ingenious invention: the indigenous Kpelle script”, *Liberian Studies Journal* 15:2, 1990.
- Winkler, E. *Kpelle-English dictionary with English-Kpelle glossary*. Bloomington: Indiana University Linguistics Club, 1997.