Reply to Lloyd B. Anderson
Ecological Linguistics
P.O. Box 15156
Washington, D.C., 20003, USA
(202) 546-5862

UTC/91-31
26 February 1991

Response to paper by Joseph Becker, dated 10 January 1991

I also have enormous respect for Joe Becker, especially for pushing forward the idea of a single integrated code including all alphabets and for much diligence and work in accumulating data needed for that effort. However:

1. Contrary to Joe Becker's claim about this, he and I do not agree even about where we disagree. I hope I would never say that I believe "there exists in an absolute sense a Single One True Divinely Mandated Encoding for every script." That would seem to me a very immodest statement if anyone should actually make it, because it appears to be claiming status as God. Because of all the very loaded terms Mr. Becker has inserted into this statement, it is difficult to discuss at all as it stands.

There are more alternatives than the extreme statement above and the statement Mr. Becker offers as his only alternative, that "the selection of code elements and the assignment of codes to them is a nearly arbitrary consensual process." That is equally I think a very wrong statement.

2. There is a very normal sense in which a coding can be right or wrong, or at least relatively more right or more wrong. The selection of code elements and their assignment to codes is not merely an arbitrary consensual process for us in coding committees.

This is actually a position of great MODESTY, that we on computer code committees are not creating scripts but accomodating the ones which actually exist. We have a special responsibility to others, not some special privelege to exercise our own unconstrained free will on codings. While special knowledge of codings specialists can suggest that some proposed coding is "unworkable", we do not have the right to declare it "unacceptable" based merely on exercise of our personal will. Our obligation is to the script and its users, to make its coding maximally "workable", as far as technical computational possibilities and our limited creative abilities permit.

3. The (partially arbitrary) consensual process to which Mr. Becker refers actually occurred long ago over hundreds or thousands of years, and its agents are the populations which actually use the script. It would be presumptuous of us to substitute our own judgements for those of generations of users. Their script already has a structure before we ever get to learning about it.

So how do we discover what the elements of a writing system are and how they are structured?

4. The question of what the functionally independent units are of an information-carrying system (sound system, writing system, or etc.) is one which has been analyzed for over a century by the well-established field of structural linguistics. The primary principles involve proofs for independence, distinctiveness, and free combinatorial ability. Some of the principles were in fact used by Mr. Becker in his paper on repertoires. In the case of Ethiopic, we can talk about a repertoire of consonants and vowels, or a repertoire of syllables. The syllables permitted are exactly those which result from combining the consonants with the 7+7 vowels, and adding two or three special additional forms.

Dealing with the specifics concerning Ethiopic.

5. I specifically stated that one could treat the Ethiopic script as either a syllabary or an alphabet (read this as meaning that "either treatment would be some sort of maximally workable coding"), but that there is evidence it is an alphabet because all combinations of consonants and vowels can occur which make phonetic sense, and no forms which are not recognizable as combinations of those consonants and those vowels can occur. (There may be minor exceptions to this, which is one reason why there is no SINGLE best coding. See also row F of the tables.)

6. The Ethiopic code table currently in Unicode (and DIS 10646) displaces from their normal order the following forms and some others.

    q̈ä q̈ii q̈ee q̈ə   k̈ä k̈ii k̈ee k̈ə   ẍä ẍii ẍee ẍə   g̈ä g̈ii g̈ee g̈ə

These are part of the most basic Ethiopic alphabet, not forms used for foreign borrowings. There is no question about their alphabetical order in sources of which I have knowledge. They are normally ordered for each consonant as additional vowels after the basic seven vowels.

    qä quu qii qaa qee qə qoo   q̈ä q̈ii q̈aa q̈ee q̈ə
    kä kuu kii kaa kee kə koo   k̈ä k̈ii k̈aa k̈ee k̈ə
    xä xuu xii xaa xee xə xoo   ẍä ẍii ẍaa ẍee ẍə
    gä guu gii gaa gee gə goo   g̈ä g̈ii g̈aa g̈ee g̈ə

As far as I can observe, there has been no reason given for departing from the standard order for these forms, except the wilful imposition of an 8-member row for each consonant. Since the 12 vowels shown here can in fact combine with most consonants, and since there are an additional two vowels combinable with at least some and perhaps with most consonants, a longer row for each consonant would be one of the most "workable" solutions. This is treating the Ethiopic script as a syllabary. The same organization would be appropriate and highly "workable" for coding these forms as "rendering glyphs" if the Ethiopic script is treated as an alphabet of the Indic type (in which the basic consonant letters contain an "inherent" vowel).

7. Mr. Becker states that there are no symbols for writing the separate vowels in Japanese Kana or in Ethiopic (except of course those used for writing "initial" vowels, those combined with a special weak initial consonant or null consonant). Agreed, in Kana there are not. But in Ethiopic, over 80% of basic forms transparently contain a recognizable form of the vowel.

The actual counts for a part of the core Amharic forms are as follows (giving a minimum value because of the unsegmentable nature of the "regular" form for the shwa voweled forms). The additional five or seven /w/-voweled syllables are even more regular than these.

Vowel	uu	ii	aa	ee	shwa	oo	Percents all vowels
First form	31	30	22	30		21	68% regular
Second form			7		10	10	14% secondary form
Irregular	2	3	4	3	23	2	19% irregular form
% Irregular	6%	9%	12%	9%	69%	6%	

8. The attempt to structure Ethiopic script in regular-length rows (of 8 or 9 or 12 or 15 or 16 members) is itself a partial recognition that the functional dimensions of the script are consonants vs. vowels. If the system were purely a syllabary (i.e. a list of under 600 forms without strongly systematic internal structure), we could simply arrange all the forms one after another with slight variation in row length.

9. The fact that distinctions of sound have been lost, and that there are three different consonant letters all pronounced [h], is not relevant to the question of distinctive and independently functioning units of the **writing** system. It concerns the relations between sounds and writing, and the difficulty of correct spelling. English has both "k" and "q", once distinct, but now merged in sound. We have two different pronunciations for the sequence "gi..." (gin, give) and two different spellings "gin" and "jin" which can be pronounced the same. This in no way changes the fact that there is a single unit "g" in the writing system, or that "k" and "q" are two distinct units.

10. Ethiopic script is and has been taught as an alphabet. The teacher at the Foreign Service Institute, US Dept. of State, has or had on his wall a list of the Ethiopic alphabet, all consonants in first-series form (with inherent vowel), displayed just as someone from India would display their alphabet. This was an Ethiopic product, not something produced here.

Ethiopic script is of course also taught as a syllabary.

11. In regards to the recent conference of the Ethiopic Scientific society in the USA, Ethiopians indicated their emphatic preference for alphabetic typing, and said that only a few professional typists who were accustomed to it wanted anything to do with the traditional Ethiopic mechanical typewriter layout.

I could have stated this more strongly in one respect: All three or four (?) DOS-based systems which Ethiopians had developed and demonstrated at that conference used the alphabetic principle for keyboarding. None even offered the traditional typewriter layout, perhaps partly because of its incompleteness — not all forms can be produced from the elements on the Ethiopic typewriter.

I completely agree with Mr. Becker that keyboarding has no direct implications for character codes. I cited the point just mentioned only as evidence for the psychological reality of the units. [A counterargument could be influence from living in an alphabetic land.] Psychological reality of the units of a writing system is indeed relevant to choosing a "workable" coding for that alphabet, because of questions of ease of use.

Of course both alphabetic and typewriter-imitating keyboard layouts need to be provided for use in Ethiopia. The point is that those who have a choice prefer the alphabetic typing, because the combinations are predictable from the units in which Ethiopians think about their script, and the users prefer not to have to choose them manually. The mechanical typewriter keyboard is a result of technology limitations, and does not directly reflect the psychologically real functioning units.

12. It is true that alphabetical sorting order in many dictionaries merges the different syllables with the same pronunciation (especially the three "h" consonant series, or the two "s" series, etc.). An XLAT "translate" table to do this in software will be simple wherever the archaic consonants are placed in the code table ordering, and simpler if rows are of constant length and structure.

13. Since the alternative codings which provide for the necessary total number of glyphs occupy (circularly) about the same amount of code table space, there is every reason to order the forms in the way which is maximally workable, reflecting as much as we can of the structure of the writing system itself rather than imposing externally invented criteria from outside.

I fail to understand how the order or arrangement of Ethiopic syllables is overconstrained, except in the single point that the archaic ordering for a few consonants is different from the modern ordering. We cannot match both (and cannot absolutely match the modern in any event). But we can certainly avoid any unnecessary deviations from ordering.

All knowledge about the normal Ethiopic ordering which I have seen points in the same direction, with the sole exception of the consonant letter /v/ and the archaic vs. modern positions for the archaic consonants pronounced / h, s /. Since two different codes cannot be ordered as identical without intervening software, we cannot satisfy the modern ordering with the code table alone. But we can satisfy the older traditional ordering, or we can approximate modern ordering. These differences could involve merely switching the orders of three or four consonantal series, retaining their row-internal ordering constant. The accompanying charts are in traditional order.

Except for the incompatibility noted between archaic and modern ordering for a very few consonants, we can have our cake and eat it too. We can list all permissible syllabic forms in a coding which is simultaneously in proper alphabetical order AND in a consistent pattern (barring the single case of three distinct variants for the syllable [fwaa]), with the same number of code cells for each consonant. Some combinations will be extremely rare, but nearly all can in principle occur.

14. The attempt to exhaustively list Ethiopic syllables as an arbitrary irregular system is an attempt to make Ethiopic into a finite repertoire on a level (syllables) where it is structurally rather an unconstrained result of combinations of consonant and vowel units, with irregularity restricted to the level of graphic rendering ligatures.

The repertoire is structurally finite on the level of consonant and vowel units; it is combinatorially generative (secondary, not limited by listing) on the level of syllables.