Problems of Orthography and Word Division

in East African Vernacular Bantu Languages[1]

H. Beck

Some Missions had already started to explore the country and eventually succeeded in finding a foothold for their work in East Africa even before the Zanzibar Treaty of 1890, which decided on the political future and the boundaries of Tanganyika. (In 1867 the Universities Mission to Central Africa began work in Usambara; the Church Missionary Society, who first went as far as Uganda, from there turned back again to Tanganyika; and in 1886 the Bethel Mission began work with a first establishment in Zanzibar.) However, a greater number of missionary societies or churches started to work in the Territory after the initial establishment of colonial rule, which did a great deal to put an end to slave trade and tribal warfare and to pacify the country. As at that time the country was under German rule, naturally a considerable number of the Churches and Missions which came to work in the area were of German origin, e.g. the Berlin and Moravian Mission (which at that time was under a German Board), and the Bethel and the Leipzig Mission whose work is now integrated in the Lutheran Churches of Tanganyika.

Except for the coastal area, where besides Arabic they found considerable Swahili literature, they came to work in a country with a great number of different vernaculars in which nothing whatsoever had been written, and where very often people did not even know that the art of reading and writing existed or what it might be good for.

Under these circumstances thorough language study by the missionaries was the prerequisite for effective missionary work. In due course they faced the task of reducing the spoken word to writing, i.e. of considering and deciding upon matters of orthography in particular, and in general on the principles which were involved in the process of developing the spoken into a written language.

A great number of missionaries who came to work here in (formerly German) East Africa had done preliminary studies at the Institute for African Languages at Hamburg University, where the extremely gifted and well known Prof. C. Meinhof did a great deal to prepare them for the linguistic tasks which were awaiting them in Africa.

As a consequence the missionaries' approach to language problems, in co-operation and consultation with their great teacher, was largely a scientific one, as we can easily observe in the older Scripture editions, some of which are still being used nowadays without any revision work having been done since their first publication.

The essential feature of their approach was the exactness with which matters of accidence and orthography were dealt, which often far exceeded the degree which we now consider to be expedient or practicable.

[1] This paper was originally read at the Bible Society Conference held at Utengule Tanganyika in 1959.

Some Characteristics of Earlier Orthographies

a. For sounds which were completely or even only slightly different from those in languages with which the missionaries were already familiar, new letters were added to the standard alphabet.

For example, the palatal nasal (without aspiration) *ń* which is similar to the *gn* in the French *ignorer*. In some Scriptures it is still used while in others it has been replaced by *ny*, although the *ń* is definitely slightly different from the nasal alveolar *n* plus *y*, i.e. semi-vowel *i* (phon. writing /i̯/). In some languages where this replacement or rather simplification has taken place a long time ago, as in Swahili, we do not notice anything strange about it. Nobody would any longer think of writing *-ńoa* for *-nyoa* (to shave), although the former is used as the phonetically correct spelling in *Bantu-Phonology* by C. Meinhof - N.v. Warmelo. Where the Nyiha[2] NT (edit. 1913) reads *ińuńi* (bird) the Inamwanga NT (edit. 1949) already reads *inyunyi*, which has about the same or at least a similar pronunciation. In languages other than Swahili, where the difference between *ń* and *ny* may be greater than in Swahili, in Nyiha for example, some of the older people may be more reluctant to accept a change of orthography on this point, but there is no doubt that the younger generation is all for it.

In the first Gospel edition (Matthew) in Nyiha of 1904 the palatal fricative voiceless sound, which we now write *sh* (*ishintu* - the thing), was still written *x*. The same Gospel edition still follows the Meinhof orthography which used an italicized letter for the voiced bilabial fricative '*v*' ('*ivintu*' - the things), which occurs rather frequently and was already replaced in the 1913 edition of the whole NT in Nyiha by an ordinary 'v'. Even the 1913 Nyiha NT already shows more simplification of the orthography, which in the 1904 edition of Matthew was still much more like the scientific orthography of Meinhof:

Meinhof	Matthew 1904	NT 1913	NT Revision 1958-60
tš	tj	c	ch
dj	dj	j	j
j	j	y	y
kh	k	k	k
ph	p	p	p
th	t	t	t
v	v	ɣ	ɣ
v	*v*	v	v
ń	ń	ń	ng'
ń	ń	ń	ny
š	x	sh	sh

The later editions of Nyiha Scriptures (see table above) show without doubt the trend of development towards further simplification of orthography which will finally allow us to write all East African vernaculars concerned on an ordinary typewriter, a matter which is essential for literacy work on an extensive scale, including processes

[2] Often listed as Nyika, as it was formerly spelt.

of duplication and printing, which will be the easier and cheaper the fewer special letters and types are needed. As long as any literacy work in the vernaculars is needed and desirable, the reproduction should be as convenient as possible. Many of them have only the NT, a catechism and a hymnbook, or even less. Although simplification of orthography is a very important point, we will see that it should not be oversimplified and done too lightly, as some pitfalls may be hidden in too hasty a procedure of formal streamlining. We will later come back to this matter.

b. Detailed differentiation was made of vowel sounds in particular, which usually would each be rendered by one type of letter only.

Plenty of good examples of this are to be found in the Nyakyusa NT, which one can hardly read without stumbling across them, and which urgently needs revision. There are many diacritical signs in it (lines and dots above and below the vowels) which most readers, even ministers and missionaries, would not be able to understand or to explain. Here are a few examples from the first chapter of John's Gospel which illustrate the point:

ālịko, fịpelīgwe, mụ, mūnụma, nīsīle etc.

The diacritical marks qualify the vowel as open or closed, as long or short:

Closed vowels:	ẹ	as *é* in French *été*
	ị	as *ee* in English *seed*
	ọ	as (nearly) *oo* in Engl. *moor*
	ụ	as *oo* in Engl. *food.*
Open vowels:	*e*	as *è* in French *zèle*
	i	as *i* in Engl. *bitter*
	o	as *o* in Engl. *ox*
	u	as *oo* in Engl. *foot.*
Long vowels:	ā, ē, ī, ō, ū.	
Short vowels:	a, e, i, o, u.	

The Nyakyusa NT obviously prefers to mark the closed and long vowels, not specially marking those which are relatively open and short.

As the Scriptures are not meant especially for linguistically trained missionaries who know these principles better than the vernacular itself, but for people who speak it as their mother tongue, all these distinctive marks of vowel quality can be omitted without harm in the process of orthographic simplification.

Problems of Orthographic Simplification

There may however be some difficulties of which we must be aware if we want to avoid trouble when we do away with diacritical marks. Sheer omission may not only render the reading and understanding, as well as the pronunciation, more difficult, but may also change the meaning of words to an inadmissable degree. Many linguists, translators and revisors will have to wrestle with particular problems in which basic requirements of orthographic simplification come up against the insufficiency of the conventional alphabet as it is used in the Indo-

European language family. It will not always be easy to find a solution and sometimes diacritical marks will have to be retained, for example:

(Nyiha) *ukuvuzya* - 'to speak to, to tell someone'
ukuṿuzya - 'to make increase into great numbers'
(causative of *-ṿula*, 'to be plentiful').

Difference in pronunciation and meaning are considerable, and both words are used in the Nyiha NT. There is the dentilabial voiced fricative *ṿ* and the bilabial voiced fricative *v*. In a recent edition of the rules and regulations of a Nyiha Co-operative Society it was attempted to render both by simple *v* only, with the result that some Nyiha people themselves got their pencils out and put the diacritical dot underneath, wherever it was needed. In this instance the revisors of the Nyiha NT have found it necessary to retain it because otherwise considerable difficulties would arise. It is, moreover, not too difficult to write the *ṿ* in two stages on the typewriter: *v* and then the dot underneath, and fortunately it occurs far less frequently than the *v* which is bilabial. To use the *w* for any one of these sounds is out of the question as it stands for the non-syllabic *u* (phon. writing /u̯/) as in Swahili. There are a few other languages (e.g. Ewe in West Africa) where the differentiation of the bilabial and the dentilabial *v* seems to be indispensable. In some instances the experiment has been tried of introducing for one of them the Greek beta, or an italic *v*, or diacritical marks resembling a Greek breathing above a *w*, but all these possibilities commend themselves less than the simple dot underneath the *v*, to which people have now become very well used in the course of past decades.

The Introduction of New Types of Letters

The introduction of new types of letters which are not familiar to children who learn to read English or Swahili in the schools, should be avoided as much as possible. Firstly, they make typing, duplicating and printing more difficult and more expensive. Secondly, they may deter people from using the Scriptures in their vernacular. Where such special types are being used, investigation should be made as to how far they could be replaced by other letters which are more generally used and can be found on an ordinary typewriter. That sounds very trivial but is important in these days when papers for wide circulation are no longer copied by hand. This consideration concerns for example the velar-nasal *ng'* (Swahili writing) which is rendered in Inamwanga by the tailed *ŋ* as in *iŋanda*.[3]

The orthography should be so simple that even those people can read it who are not very educated and who can only read simple matter, yet it should be sufficiently specific to exclude serious misunderstandings and ambiguities. It will not always be easy to find the mean which is satisfactory for readers of all levels of education. But consultation and conferences in areas with many different vernaculars

[3] In recent Inamwanga literature this letter has indeed been replaced by *ng'* (this was reported during the Untengule Bible Soc. Conference after this paper was read).

which may partly overlap will be useful, and they will help those who work in any one of them to see the requirements, difficulties and possibilities clearer in their particular vernacular, and how their work can best be adapted to the generally prevailing trend of development.

Word Division

The question of right and practicable word division should not be decided on a purely traditional or experimental basis, but with knowledge of and on the ground of the general structure of the Bantu languages. Even for those who speak the language concerned as their mother tongue, it will not always be easy to decide on matters of word separation unless they have a comprehensive knowledge of the structure of the language group to which it belongs.

One of the obvious laws, for example, which seems to govern many of the East African Bantu languages (although it does not so much apply to Swahili where connection of small words with longer ones is less frequent, and where a comparatively large number of nouns begin with a consonant) is that small conjunctions are often attracted to the following noun:

(Nyiha)	*nishivanza*	for	*na ishivanza*	(and the church)
	numuntu	for	*na umuntu*	(and the man)
	namagunda	for	*na amagunda*	(and the fields)

Here the separation in the Nyamwanga NT must appear as artificial and inconsistent when it reads:

ni vintu (Mk. 7:13) for *nivintu*; actually *na ivintu* (and the things)
na makasa (Mk. 7:5) for *namakasa*; actually *na amakasa* (and the hands)
nu mulilo (Mk. 9:44) for *numulilo*; actually *na umulilo* (and the fire)

words which organically belong together and which thus contracted should not be divided up but rather be left together in the way they fit together and are read together, the first one having dropped its final vowel in consequence of the initial one of the noun to which it was attracted.

We will have to be careful that words do not become confusingly long, but we should on the other hand be free from the misunderstanding that a multitude of partial words, some of them not amounting to more than one letter, are easier to read than fewer words which are a little longer. It is a generally made observation, which is not ignored in Primary Education, that even in the first stages of learning the pupil is apt to read not letter by letter, but rather by catching with the eye the picture of the whole word as it appears in print, especially if it is only short. The advanced reader will be able to read the "word-picture" of even longer words and parts of a complete sentence without looking consciously at the individual letters which make them up.

What we have said concerning the small copula *na* is also relevant for the locatives as they are used in many of the Bantu languages:

(Nyiha)	*mwivoti* (Jn. 6:17)	for	*mu ivoti*	(in the boat)
	mwilisyo (Lk. 8:32)	for	*mu ilisyo*	(in the pastures)
	pigamba (Lk. 8:32)	for	*pa igamba*	(on the mountain)

but also: *munsi* for *mu insi* (in the country)
munkaya for *mu inkaya* (in the village)
mumbombo for *mu imbombo* (in at-work)

Besides the simple locative which is prefixed to the noun there is another type which is constructed with an added *a* (similar to the way the genitive pronouns are made up):

kwa for *ku* plus *a*
mwa for *mu* plus *a*

and which we may call a 'heavy type locative' which is used in some instances and has so much weight, that it remains by itself, e.g.:

mwa sumbi (in the sea).

In some of the Bantu languages, Swahili for example, the locative has so much weight that the following possessive pronoun is not constructed in congruency with the class proper to which it belongs, but rather according to the locative class to which it has then been transferred:

nyumba yangu, (Swahili: my house)
but *nyumbani mwangu,* (in my house).

Even where this is not the case (as in Nyiha, for example):

ivoti yane (my boat)
and *mwivoti yane* (in my boat),

the examples quoted indicate that the locative in its basic form is connected to the noun with which it goes, while this is not fully the case with its relative form in which it often precedes the verbal form:

munkaya mw'akuliye (Mk. 6:1: into the town in which he had grown up; into his own country).

This relative form of the locative can however also precede a noun and then even more attention is required:

munsi ya mu-Galilayi, mw'ivoti lihasulila kushipuma (Nyiha Mk. 6:53: in the land of Galilee *where* the boat pushed on the shore; i.e. they drew ashore).

Note the difference:

mwivoti - in the boat (simple locative)
mw'ivoti . . . - where the boat . . . (relative form of the locative).

This instance shows how carefully each term and expression needs to be considered and how great the danger of error can be when matters are handled more formally than with real understanding.

In the Nyamwanga NT the locative prefixes, too, are separated from the nouns to which they obviously belong:

mwi yulu, from *mu iyulu* (Mk. 10:21) - (in heaven)
pi sambilizyo, from *pa isambilizyo* (Mk. 11:18) - (at the teaching)
mwi tempeli, from *mu itempeli* (frequently used) - (in the temple)
mu wumi, from *mu uwumi,* which should probably be written *muwumi* (in the life)
mu wufumu, from *mu uwufumu,* which should probably be written *muwufumu* (in the kingdom),

but also:

mu nsi, from *mu insi* (in the country)
mu nkaya, from *mu inkaya* (in the village)
mu ŋanda, from *mu iŋanda* (in the house),

which locatives should probably be written as one word only: *mwiyulu, pisambilizyo, mwitempeli, muwumi, muwufumu, munsi, munkaya, muŋanda* etc. Our problem has here been reversed; while we are dealing with word division we must also be concerned with the right junction of words.

The Use of Capital Letters

In many instances where the first letter of a word is automatically written with a capital letter, the writers may just have followed a habit to which we are used from the Indo-European languages, without having given sufficient thought to the particular structure of the Bantu languages. The Indo-European languages have no class prefixes which correspond to those of the Bantu languages, whilst the Bantu languages have not the definite articles of the Indo-European languages. These essential differences should not be by-passed in matters of orthography, which includes the question of how to use capital letters.

In the English language it makes little difference to the word-picture when we write: Jew, Jews, Jewish. The first basic part of the word always remains the same and as such attracts the eye, which then in due course also notes the form in which it appears, singular and plural, for example, only being indicated by the omitted or added final *s*.

However, it makes much more difference to the total word-picture (which is important for the whole process of reading) if the eye-catching initial letter, written as a capital, changes each time with the form in which the name proper is used:

Myahudi, Wayahudi, Kiyahudi (Swahili: Jew, Jews and Jewish), where the original part, which is a derivation of a proper name:

Ioudaios, Ioudaioi and *Ioudaikos* (Greek),

in Swahili for example is not so strikingly obvious as could and should be the case if it had been written as:

mYahudi, waYahudi, and *kiYahudi,*

the prefixes being only constructive accessories to the basic part and core of the word, which as such remains unchanged.

(In addition, in the above instance, a 4-syllabic word: *M - ya - hu - di* is virtually, as far as the orthography is concerned, changed into a 3-syllabic word, because in the Bantu languages *y* is almost generally used for the semi-vowel *i* (phon. writing /i̯/) which has not enough weight to constitute a syllable: *mya* /mi̯a/ is one syllable, whilst *mia* /mi-a/ has two syllables. The *Standard Swahili-English Dictionary* (Oxford University Press, 1939) contains only four other cases where *y* after the class prefix *m* (including the plural forms they would belong to the word-classes 1 to 4: *m-, wa-, m-, mi-*) must be considered and treated as initial of a syllable, following a consonant in the singular and a vowel in the plural form.)

A number of Scriptures in various vernaculars follow the principle of writing the name proper with a capital although it is preceded by the class-prefixes:

In the Nyamwanga NT this policy prevails, although there is a great deal of inconsistency:

muYuda	(Rom. 2:17)
awaYuda	(Jn. 19:31)
waHeleni	(Acts 9:29)
awaFalisi	(Mk. 7:1)
waHelodi	(Mk. 8:18)
maSabata	(Acts 13:27); but: *Ipasaka* (Jn. 18:28)
ciHebele	(Jn. 19:20) likewise: *ciRoma, ciGriki*
awaYisraeli	(Acts 13:24) but: *a wayisraeli* (Rom. 9:4)
	Awayisraeli (Rom. 9:6)
	aWayisraeli (Rom. 10:19)
	Awa Yuda (Jn. 18:31)

The Mambwe-Lungu NT reads:

aYudea (Acts 2:5)
aPatia, aMedia etc. (Acts 2:9)
muYudea (Rom. 3:1)
kiYeblai, kiLoma, kiYela (Jn. 19:20).

And the same principles should certainly also be applied to nouns where they are used as names of honour and are therefore capitalized. Here, however, as there is usually no change from singular to plural but only the change from the nominative form (with prefixed vowel) to the predicative form (with simple class-prefix), the unchanging part of the word (of which the root without prefix is never used by itself) should be capitalized. For example in Nyiha:

Instead of *Umwana* or *u Mwana* (Son),
write *uMwana* (in the middle of a sentence),
or *UMwana* (at the beginning of a sentence).
(The predicative form *Mwana* remains unchanged.)

Likewise instead of *u Mulungu* or *Umulungu* (God),
write *uMulungu* or *UMulungu.*
(The predicative/vocative form *Mulungu* remains unchanged.)

In many vernaculars it would be even more important than in Swahili to capitalize the unchanging word centre, as they not only have changing simple class-prefixes, but in the nominative case, for example, also prefix the vowel of the prefix proper (*mu-, umu-, va-,* and *ava-*), so that a name which is used with such prefixes might have four different initial capital letters if the first one is always capitalized, whilst it would always be the same if the central name is capitalized (*umuYuda, muYuda, avaYuda, vaYuda* instead of: *Umuyuda, Muyuda, Avayuda,* and *Vayuda*). These secondary vowel prefixes, taken over from the first word-class (*m-tu* in Swahili, but in Nyiha for example: *mu-ntu* and *umu-ntu*) are also used with proper names which, referring to people, belong essentially to the same class of words:

UYesu, uYesu (one word!) instead of *Uyesu* or *u Yesu;*
UMwana, uMwana instead of *Umwana, u Mwana* or *U Mwana.*

If the first letter is always capitalized without consideration of the word-structure there would be many more changes in the word-pictures in other vernaculars than there are in the Swahili language. In the way indicated they can easily be avoided. As long as the orthography of the languages is still so much in the process of formation and re-formation from edition to edition of Scripture (in Swahili this process came to a halt, for the time being, when the present Standard Swahili was agreed upon) it is possible and admissible, if not even necessary, to give these matters due consideration, in order that the most suitable form may be found, which for many vernaculars may not necessarily be that of Swahili. It should, however, be recommended that such problems as those which in this paper could only be broadly indicated, be more closely examined and studied by area committees of linguistic experts, who could make suggestions to effect a higher degree of co-ordination and congruency in language policy for the Scriptures to be published by the Bible Societies in general, and by the BFBS in particular. In such ways, also, more guidance could be provided for the development of adequate orthography.

Literature

1. *Bantu-Phonology,* C. Meinhof - N. v. Warmelo, (Dietrich Reimer/ Ernst Vohsen, Berlin, 1932).
2. *Grundzüge einer vergleichenden Grammatik der Bantusprachen,* C. Meinhof, (Verlag von Eckhardt & Messtorff, Hamburg, 1948).
3. *Lautlehre und Grammatik des Nyiha,* Dr. J. Busse, Hamburg, 1943 (is being printed; for the above article only a typewritten copy could be used).

Scripture editions referred to:

1. Gospel of Matthew in Nyiha (Nyika), printed in Berlin 1904.
2. The New Testament in Nyasa Nyiha (Nyika). First edition 1913 (BFBS), reprints in 1949 and 1955.
3. The New Testament in Konde. BFBS and NBSS, 1954.
4. The New Testament in Chinamwanga. NBSS, 1949.
5. The New Testament in Mambwe-Lungu. BFBS, 1950.

HELPS FOR TRANSLATORS

The third volume in the B.F.B.S. Greek-English Diglot series, the *Gospel of John,* has now appeared. Also, the *Indian Word List,* already available to translators in India, will now be granted under certain circumstances to workers in other fields, if it is likely to be of real value to them.

Old Testament Translation Problems and the *Translator's Commentary on Mark* which are now being printed are due to appear shortly, and requests may be sent to the appropriate Bible Society. Further details will be found as usual on the last page.