THE USE OF NON-ROMAN SCRIPT FOR NEW LANGUAGES

WILLIAM A. SMALLEY

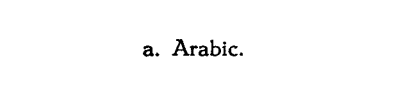
In many parts of the world the predominant language is not written in Roman script but in some other tradition such as Arabic, Chinese or Devanagari. In such cases the question arises as to which script should be used for writing tribal languages which have been previously unwritten. In the following article Dr. Smalley discusses this important subject in a general way. It is the first part of a longer article which will be printed in full in a forthcoming volume in the U.B.S. series "Helps for Translators" entitled, Orthography Studies: Articles on New Writing Systems, *which will largely consist of reprints of articles on orthography which have appeared in* The Bible Translator. *It is hoped to publish this book in 1963. Ed.*

In modern times, as writing systems have been prepared for previously unwritten languages of Africa, Asia, North and South America, and Oceanea, these have usually been based on some form of the "Roman" alphabet, as the general Western Europe alphabet tradition is called. A notable exception, of course, has been within the Soviet Union, where the Russian or Cyrillic alphabet has been used for large numbers of languages, particularly of Asia.

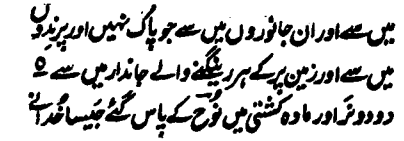
The reasons for this modern spread of the Roman alphabet are easy to see. Through the colonial period the agents of change, of education, of dissemination of literature, were principally Western Europeans or Americans. Because the technology of printing got its start and greatest development in the West it was adapted primarily to Roman script. Other fonts of type were hard to obtain and often expensive, if not nonexistent. From the printing point of view the Roman alphabet has the advantage over some alphabets in the small number of characters, and the linearity of the system. Vowels and consonants follow each other on the line, whereas in some systems such as the Thai (see illustration e on p. 202), there are arrangements of vowels and other signs as diacritics in an "orbit" around the consonant, which makes composition of type somewhat more difficult. Arabic presents similar difficulties.

Most Westerners who prepare new orthographies for peoples without a writing tradition seem to feel that the Roman system is easier to teach to nonliterates. There may be a certain amount of ethnocentrism here, of reacting to one's own way as the easiest way of doing things, but it would seem that there is objective validity as well, so far as some of the non-Roman systems are concerned. The Thai system for example, at its best, is systemically more intricate, more difficult to remember, than is the Roman system, when each is carefully and consistently applied.

In earlier centuries, however, two other great orthographic traditions which are still very widely used spread with culture and



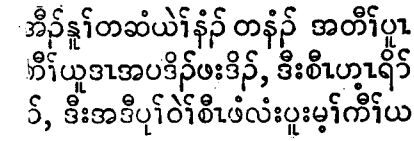
a. Arabic.



b. Urdu, *using the Persian form of Arabic script.*

च मेरी जो वाचा तुम्हारे और सब जीवित
ाणियों के साथ बंधी है ; उस को मैं स्मरण
गा जलप्रलय फिर न होगा जिस से सब

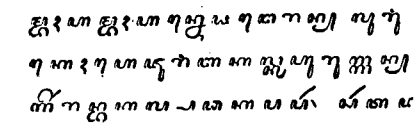
c. Hindi, *using the Devanagari script.*



d. Karen: Sgaw, *using Burmese script, derived from India.*

จะมีโทษถึงไฟนรก. เหตุฉะนั้น
แท่นแล้ว, และระลึกขึ้นได้ว่าท่าน

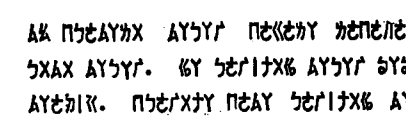
e. Thai. *Script derived from India, via Cambodian.*



f. Javanese. *Script derived from India.*

የዛብሎን ፡ ምድርና ፡ የንፍታሌም ፡ ምድር ፡
የባሕር ፡ መንገድ ፡ በዮርዳኖስ ፡ ማዶ ፡ የአ
ሕዛብ ፡ ገሊላ ፤ በጨለማ ፡ የተቀመጠው ፡

g. Amharic.



h. Sediq, *using the Chinese Phonetic script.*

SAMPLES OF NON-ROMAN ORTHOGRAPHY

Included here are samples (taken from Bible translations) of some of the scripts to which reference is made in this paper. For some of these languages other scripts are in common use, and in the case of Javanese the Roman script is more important today.

religion from language to language.[1] These were the Arabic tradition, and that of the scripts related to the Devanagari[2] of India. The Chinese tradition was extended as well, to the Japanese language, to Vietnamese (which later switched over to Roman alphabet), to some of the non-Chinese languages of China, plus some influence on Korean. The Chinese spread would doubtless have been much greater if it were not for the essential inefficiency of the Chinese system, which requires the memorization of hundreds of logographs (individual characters which do not represent sounds, but words, morphemes, or syllables which are parts of particular words or morphemes. The present Chinese Communist emphasis on orthography reform is motivated by a desire to change from this educationally cumbersome system to a more easily learnable orthography.

Arabic script

Next to the Roman alphabet, it is the Arabic system which has spread most widely, as it has followed the expansion of Islam. Arabic writing systems have been used for languages as remote (both geographically and genetically) as Spanish and Malay. The languages of the Middle East,[3] the principal languages of North Africa, Urdu,

---

[1] The scope of this paper extends far beyond my specialized knowledge. The information, furthermore, is not always available in published sources. The original data on the adaptation of non-Roman scripts to the various languages were obtained in detail from the individuals acknowledged in the appropriate places. The first draft of this paper was later circulated to these same people and to a number of others familiar with the problem of the missionary use of non-Roman scripts in one part of the world or another. Almost everyone replied. To all of these people I am greatly indebted for suggestions that have improved the paper or corrected it. The major contributions are acknowledged in footnotes, but scores of minor improvements must be acknowledged by a simple listing of names: Lois Callaway (Thailand), Ralph R. Covell (Taiwan), Wesley J. Culshaw (India), H. A. Gleason, Jr. (Arabic, and languages of India), Herbert G. Grether (Thailand), E. E. Heimbach (Thailand), Harvey T. Hoekstra (Sudan), Paul E. Hostetter (Sudan and Pakistan), Edwin J. Hudspith (Thailand), Beulah Johnston (Thailand), James B. Keefer (Ethiopia), Geoffrey E. Marrison (India), Gene Nelson (Thailand), Eugene A. Nida, Herbert C. Purnell (Thailand), Donald N. Rulison (Thailand), Donald Schlatter (Thailand), J. L. Swellengrebel (Indonesia). These people are, of course, in no way responsible for mistakes of fact or interpretation which may still remain.

[2] The Devanagari script is the one employed for Sanscrit and for Hindi and Marathi. There is some debate as to whether the related scripts are derived from Devanagari or from some common source.

[3] Turkish as used in Turkey is a notable exception now. It is in modern times, however, that under the decrees of Kemal Pasha Ataturk Turkish was changed to Roman script. Turkish and various languages and dialects related to it have been written at various times in at least the following writing systems: Cyrillic, Armenian, Greek, Hebrew, Arabic, and Roman. See David Diringer, *The Alphabet, A Key to the History of Mankind* (London: Hutchinson's Scientific and Technical Publications, 1953), pp. 567-568. I am indebted to Diringer's book for some of the facts contained in the first part of this article. His discussion of alphabets and their spread is a fascinating one, but the book should be used with caution as it is quite naïve about language structure and phonology, and about language relationships at some points.

Kashmiri, Sindi, Pashto and some other languages in Pakistan, as well as Malay in Indonesia, have traditionally been written in Arabic script, and most of these important world languages still are, although there was a shift to Roman during Dutch rule in Indonesia. Javanese was written in Arabic script only for religious purposes, and other languages in Indonesia have made some use of Arabic scripts alongside ones derived from India.

There are differences from language to language within the Arabic writing tradition. Shapes of letters vary from region to region. Differences in the number of phonemes of different languages have brought about modification of the number of characters, changes in their value, etc., so that the various orthographies in the Arabic tradition differ in ways somewhat analogous to such differences in the Roman tradition as English, German with its Gothic letters, Twi with such characters as ɔ, ɛ, ŋ, French with è, é, and Vietnamese with ư, ȯ.

However, the basic conventions are remarkably uniform throughout the area. Additional consonants number perhaps a dozen. No letter is dropped since Arabic loan words are almost universally spelled as they are in Arabic. Some Roman systems use as many additional letters as this but they almost all depart more widely in some of the values which they assign the letters. Some languages written in Roman script use digraphs or trigraphs (like English *ng* for the phoneme /ŋ/ or *tch* in *fetch* for /č/) to which nothing comparable is found in the Arabic orthography tradition. [4]

The Arabic writing system has one particular characteristic which greatly facilitates adaptation from language to language. It distinguishes some of its consonants by varying numbers of dots associated with a basic symbol, above or below it. By adding and subtracting dots new symbols are easily provided. In Sindi, for example, a number of new characters have been created by writing four dots above or below a base symbol already in existence. The adaptations described for Anuak in this paper have made use of this tradition in Arabic.

On the other hand, the Arabic writing system has not been extensively modified in the use of vowels. Arabic has three vowel diacritics, and as Arabic writing had spread there has been a marked resistance to adding to this number. There has likewise been resistance to adding to the three letters used to mark long vowels or diphthongs.

Devanagari script

The next most extensive writing tradition is the Indian one. Here the graphic differences are even greater than the Arabic, both within

[4] I am indebted to Prof. H. A. Gleason of the Hartford Seminary foundation for the points in this paragraph, and other comments concerning the Arabic tradition.

India itself, and in comparison with the writing system as applied to Thai, Burmese, and languages of Indonesia. [5] Modern Thai and modern Cambodian scripts do not look much alike and have had differing influences upon them, but both are ultimately derived from the ancient Khmer writing of the Cambodian civilization which in turn was modified from a script in South India. This South Indian script was very different from, but related to, the classical Devanagari script. The extension of the South Indian script in various modified forms into Southeast Asia, including Indonesia, went along with the spread of Hinduism and Buddhism. Where Islam followed into Indonesia there was occasionally an ambivalence (as in Javanese) between the Arabic system and that derived from India.

Missionaries and non-Roman scripts

When using languages in which non-Roman writing systems were already in existence, missionaries have usually learned the system and adopted it in their work without serious question. *The Book of a Thousand Tongues* [6] shows scores of reproductions of Scripture translations in some sixty different scripts. Most of these were used as they were found.

Occasionally missionaries have invented new scripts for peoples without any writing system, or one which for some reason the missionary did not want to use. James Evans, for example, devised the Cree syllabary in 1841, and S. Pollard the Pollard Script used for Miao languages and others in South China.

Occasionally missionaries have used an Arabic script as a second script, as in the Maninka language of West Africa, but have not emphasized it and it has fallen into disuse because the Roman was easier for them. Occasionally a language has been written in a non-Roman script but because of lack of evangelistic effort, absence of any church, discontinuance of the work, it has never really been seriously used. Such is the case of the Northern Thai script printing of the

---

[5] When two *languages* are related they are descended from the same parent language as *speech systems*. They may or may not have the same orthographic tradition. It is perfectly possible for the same script tradition to be used for unrelated languages (English and Vietnamese), or for two related languages to have different script traditions. In fact, Urdu and Hindi are almost identical in their language structure (though differing widely in vocabulary), but the one uses the Persian script from the Arabic tradition, and the other uses Devanagari, and other related systems for its writing. Some missionaries have had the idea that because a language like Cambodian uses a *script* related to the Devanagari (Sanscrit) script, the *language* is derived from Sanscrit. Actually the languages are of entirely different families. Sanscrit is distantly related to English, being part of the Indo-European language family. Cambodian belongs to the Mon-Khmer family.

[6] Edited by Eric M. North, published for the American Bible Society by Harper and Brothers, 1938.

Gospel of Mark in Khmu? or an earlier attempt to put Yao into Thai orthography. [7]

There are very few clear cases where missionaries adapted an existing, traditional, non-Roman alphabet to the language of minority groups well enough, and backed by enough education, for it to become well established, and where they did this sufficiently long ago for us to know that there is no doubt of its success.

One notable case is the adaptation of Burmese script to the Karen languages of Burma, first done by Jonathan Wade of the American Baptist Missionary Union in 1832. Thousands of Karen readers have used this system.

Another instance is that of the Mundari language of India. A writing system was prepared by missionaries in the latter half of the 19th century. The script employed was Devanagari, and it is well established. Devanagari has also been used for certain other tribal languages in Central India, including Gondi, Bhili, and Kurku, but not as successfully. In the latter case, for example, the lack of success was not due to any inherent difficulty in the adaptation, but to the general lack of continuity in the over-all missionary work in the area, and to the fact that non-mission efforts were directed at studying Hindi and Marathi.

The first attempt to write Santali (India) was made by the Rev. Jeremiah Phillips of the American Freewill Baptist Missionary Society, who employed Bengali script. Later the larger missions among the Santali adopted Roman, but this has never met the total need. Currently Scriptures are circulated in Roman, Bengali, Orija and Devanagari scripts! The Santals are the largest tribal community in India and are spread out over several states, which accounts for some of the diversity of orthography. [8]

In Assam such minor languages as Boro, Garo, Khasi, Manipuri and Rabha have used Assamese script at one time or another. The first three of these, however, have tended to drop the Assamese in favor of Roman script because the non-Hindu hill peoples tend to reject identification with the Hindu Assamese. [9]

In addition to these examples from Burma and from India, there are a few cases where missionaries in the past have adapted Arabic script to minority languages, but possibly none where these have really

---

[7] This attempt used the Thai characters, but not the Thai system. It ran the letters along in linear order more like the Roman system, and unlike the Thai handling of vowels.

[8] This information on India was supplied by Dr. H. A. Gleason, and by the Rev. Wesley J. Culshaw, Translations Secretary for the Bible Society of India and Ceylon.

[9] Information from Geoffrey E. Marrison, linguist and consultant in Assam for the Bible Society of India and Ceylon.

become established in any significant way. Unquestionably the real missionary contribution to the spread of reading and writing to new languages has been in the Roman tradition.

Missionaries are often offended by the accusation of cultural imperialism, but I think that in this lack of more serious and widespread adaptation to local orthographic traditions in Asia and North Africa there is clear documentation of the accusation. On the other hand, some missionaries, if left to themselves, would have experimented with local scripts, but were prevented by colonial administrations or other influential forces. Others were discouraged by the fear (sometimes real, sometimes imagined) that the phonemes of the language could not be adequately represented.

Certainly the difficulties which non-Roman orthographies sometimes present to letterpress printing have contributed to reluctance to their use for new languages. W. Norman Brown describes script reform which has been taking place in India, Pakistan, and Ceylon because of the requirements of modern printing, and gives an indication of the complexities of printing with Devanagari type: [10]

> A reasonably complete font of Devanagari type for Hindi requires four cases to accommodate it (against the two used for the Roman script for English); including the symbols for the numerals, punctuation, and commercial signs and a few others, it has about 500 different symbols (against roughly 230 in an English font, which latter includes such modifications of the basic lower case symbols as large and small Roman capitals, and italics, all of which have no corresponding modifications in the Devanagari). Such a Devanagari font, however, is still incomplete; a complete one may require twice as many symbols.

Current missionary experiments in the adapting of non-Roman orthographies to new languages are too young for us to assess their permanence. They seem to be most active in India, Pakistan, Sudan, Ethiopia, Thailand, and Taiwan. There is an uneasy feeling on the part of some informed people that more study should be given to the problem in many more areas under Muslim dominance in Africa, and in other such areas where non-Roman writing traditions prevail.

It must be recognized, of course, that a reverse cultural imperialism is also possible. Where Roman scripts have taken hold, as in such important tribes as the Garo, Khasis, Lushais and other Kuki tribes, as well as the Naga tribes in Assam, these peoples would consider any attempt to introduce or impose Indian scripts as a manifestation of cultural imperialism. The government of India recognizes this fact.

---

[10] W. Norman Brown, "Script Reform in Modern India, Pakistan, and Ceylon," *Journal of the American Oriental Society*, Vol. 73, No. 1 (Jan.-March 1953), p. 2.

Influence of government policy

Sometimes the motivation for missionary interest in non-Roman orthographies is stark government pressure. Today this is especially true in Taiwan, Sudan, and Ethiopia. In Taiwan the Nationalist Chinese government is opposed to Romanized systems partly because the Chinese Communists are advocating them. Even the Amoy language, which has been written in Roman orthography for over one hundred years, and in which the Bible was first published in 1884, has been threatened by this government attitude.[11] The Taiwan activity does not quite fit the main focus of our discussion in this paper. It is not the standard Chinese system of characters which is being used in the experiments under discussion, but "Chinese phonetic." This is a system used in some Chinese textbooks to give the Mandarin pronunciation alongside the standard logograph (Chinese character) because the logograph has no clear, systematic, easily learned clue as to its pronunciation. It is this "phonetic" which is being used to write Malayo-Polynesian minority languages, without the standard Chinese script. It is not a Roman system, and was not invented by the missionaries producing scripts in the Taiwanese languages. They have adapted it to some hill languages of Taiwan so that they could have a system which conforms to present government attitudes and which is used in all Chinese schools to teach the logograph for the first year.

In Sudan the government has not yet ruled finally that tribal language work must be in Arabic script, but there are strong tendencies in this direction. The American Mission (Presbyterian and Reformed) has therefore been experimenting in its use in Anuak, Murle, Nuer, etc. The translation of I John into Anuak has been published in Arabic and Roman script. The British and Foreign Bible Society has Dr. R. C. Stevenson in the area specifically studying the problem and giving guidance to the missions. The Sudan government also has an Egyptian linguist working on it.[12]

In Ethiopia, government policy has been that missionary work in the vernaculars could be carried out only in oral fashion. There was to be no vernacular literature. This halted earlier attempts to do literacy work among the Anuaks on the Ethiopian side of the border. Eventually, however, special permission was given for writing Anuak in the Amharic writing system, since the purpose of the government position is to contribute to national unification of language. The

[11] Much of the information on Taiwan in this article was supplied by the Rev. Ralph R. Covell of the Conservative Baptist Foreign Mission Society. See also John Whitehorn, "Some Language Problems in Formosa," *The Bible Translator,* Vol. 7, No. 1 (January, 1956), pp. 17-21.

[12] Information on Sudan from ABS files, especially correspondence with the Rev. Harvey Hoekstra of the American Mission, and conversations with the Rev. Paul Hostetter, now of The Literacy Center, Gujranwala, Pakistan.

adaptation of the Amharic orthography was worked out by the Rev. Milton Fisher, Presbyterian missionary, and the Gospel according to John has been published by the American Bible Society. [13]

My own interest in this problem was first aroused when government officials in Laos expressed opposition to education and literacy in minority mountain languages. We felt that perhaps government reaction might be less severe if the languages were written in Lao script.

But government pressure is not the only reason for the present experimentation in non-Roman scripts. In West Africa many Muslims can read Arabic, after a fashion. Often they do not understand the Arabic language, but they have learned to read the sounds mechanically as a religious practice. The task of transferring this reading skill to their own languages written with Arabic symbols is therefore not as great, and many can read their own language for the first time after a few minutes of effort. Years ago the Christian and Missionary Alliance in Guinea published Scripture portions in the Maninka language, using Arabic script. Recently a missionary in Liberia reports that Maninka traders coming there will buy these portions and read them, whereas they would not and could not read the Romanized Maninka writing in use among the Christians in Guinea.

In Thailand, although government policy calls for linguistic unification of the country, there is no government pressure on the missionaries using the minority languages of the country. For those people themselves, however, there are many advantages to their learning to read their own language as written in Thai script rather than in Roman script. Those of their children who get any education will get it in Thai. Government schools, although still far from reaching most of the remote areas, are spreading and opportunities for education are increasing. Those who can speak Thai will have a bridge to reading Thai. Those who learn to read Thai will quickly be able to adapt to reading in their own language, rather than having to learn an entirely new system. [14]

Present experiments

Missionaries who started out without ever introducing a Roman orthography feel the more certain now that this was the right course to take. Those who introduced a Romanized system first, find it more difficult to change. There are more problems for them, with people

[13] Data on Ethiopia and the transcription of Anuak in Amharic script were supplied by James B. Keefer of the Presbyterian Mission.

[14] See Eugene A. Nida, "Approaching Reading Through the Native Language," *Language Learning*, Vol. 2, No. 1, (Jan.-March, 1959), pp. 16-20; *The Use of Vernacular Languages in Education*, Monographs on Fundamental Education - VIII, (Paris: UNESCO, 1953), especially pp. 45-75. A limited number of reprints of Nida's article are available from the Translations Department of the American Bible Society.

who already know how to read, literature to be re-transcribed, etc. In the case of Meo there is the additional complication that the Thai-based script would not be usable in Laos, where there are many Meo Protestant Christians. In addition, Roman Catholic missionaries in Laos are publishing extensively in essentially the same Romanized script as that used in Thailand. One reason for hesitation on the part of many missionaries is that the Thai system is somewhat more cumbersome to teach, somewhat less consistent than their Roman system—even though each is worked out with the same care for consistency and linguistic equivalence.

The reception of the present experiments by the people they are designed to help in Sudan, Ethiopia, Taiwan and Thailand seems to follow some fairly consistent patterns. For people who have not been exposed to the writing of their language in Roman characters or are not educated in Western schools, the transcription based on the national tradition is usually assumed to be the natural thing. This is the direction in which they experiment themselves if they get the idea of doing so.

Unless there is strong nationalist sentiment [15] in the group, people who have been exposed to a Roman system quite often prefer it, especially at first. They are used to it. A change requires an adjustment. Sometimes minority peoples do not want to be identified with the national, majority group. In those areas, however, where the non-Roman alphabet has had to be used anyhow because of government pressure, the opposition shows signs of disappearing as the non-Roman gains ground. [16]

---

[15] Of course, strong nationalism does not always identify itself with non-Roman script. As Dr. J. L. Swellengrebel of the Netherlands Bible Society points out (in a private communication):

> In Indonesia it was certainly under Western (colonial or missionary), influence that the Arabic, Javanese, Batak, and Lampong scripts were supplanted by Roman script. The strong nationalist movement in colonial times never challenged this change in scripts. The present government, which certainly does not have pro-Western tendencies, has continued and promoted the use of Roman script in all cases. Only in S. W. Celebes (Sulawesi) there may be some nationalistic or provincialistic preference for the use of Makasarese script, in the very few books printed in this language. In Bali, with its strong and growing feeling of cultural distinctiveness, the Balinese form of Javanese script is still in use, but Roman script is on the increase since colonial times, certainly in printing, but also in writing.

[16] In Lawa an experimental change from Roman to non-Roman script was made in 1961. The Rev. Donald Schlatter now reports: "So far we are quite happy with the results of the Thai orthography out here. We will have in print a set of primers and are mimeographing Mark at the present time. We have also done a few other things for experimentation. The readers of the Roman script are learning the new although some of the older folks dread the change and have not tried to learn yet. With younger people we have the added incentive that it is a direct key to learning Thai. Those Lawa who can already read Thai have very little trouble with the new transcription, and can read in a very short time. They are slow, but it is probable that their Thai reading is also slow."

In the second part of this paper [17] we will attempt to document in brief detail the problems and feasibility of a Thai-based orthography for five minority languages of Thailand. This will be followed with brief notes from Sudan and Ethiopia. There will then be a summary of the practicability of such systems, and of the principles of adaptation which emerge, in the hope that with this much documentation, people who are faced with choices concerning non-Roman orthographies will have more information on which to base their decisions.

AGATU AND GREEK

We were concerned to find an adequate word for the term "hypocrite". The obvious definition is someone who pretends to be someone he isn't. We have found that the best way to find an equivalent is to create a situation and then ask what word is used to describe the thing or action in mind. In this instance, the first reply was, "A madman?" After giving another illustration, our informant came up with "Oh do you mean a proud person (that is, he values himself too highly)?" or, "Perhaps a deceiver?"

Now, we did not wish to get caught up in this last suggestion because Agatu has 7 terms showing 7 different ways of deceiving—and yet, one of these, *injei* (a person who deceives by bad example), was not far off the mark. Until we could find something better, it would have to do for the time being.

Then Young's Concordance reminded us that the Greek word had the thought of an actor—one who was playing the part—and on looking up Vine's Expository Dictionary we found, "it was a custom for Greek and Roman actors to speak in large masks with mechanical devices for augmenting the force of the voice; hence the word became used metaphorically of a dissembler"—and why, that was just exactly like the native *alede,* a masked 'spirit' with a reed in his mouth to disguise his true voice! We even had a photograph of one handy, and showing it to our informant Ɔmaye, he immediately said, "Oh yes, that's an ɔcɛ *gigbefu,*" meaning a disguised person acting a part—and so we found our Agatu word for hypocrite!

When applied to the Jews in Mark 7:6, "This people honours Me with their lips, but their heart is far from me...", Ɔmaye exclaimed, "Yes, ɔcɛ *gigbefu* has caught them exactly!" It may be some time before some professing "Christians" begin to realise that it has caught them too—but we feel sure that one day it will!

(Extract from a letter from Mr. & Mrs. Mackay, Agatu translators in Idah, N. Nigeria)

[17] To be printed in *Orthography Studies,* U.B.S. "Helps for Translators" series, Vol. V.