DESIGNING ORTHOGRAPHY FOR THE HA LANGUAGE

Lotta Harjula

Ha is one of the 127 living languages of Tanzania (Gordon 2005). Ha speakers form the majority of the population in the three districts of Kigoma, Kasulu, and Kibondo in the Kigoma region east and north of Lake Tanganyika. Nowadays, there are quite a number of Has also living in other parts of Tanzania, as well as in Uganda, Kenya, Rwanda, Burundi, the Democratic Republic of Congo, and Zambia.

The Ethnologue (Gordon 2005) lists the number of Ha speakers as 990,000. According to the census of 1988, the population of the Kigoma region was ca. 860,000, but by 2002 it had more than doubled to ca. 1,680,000.[1] The increase in the number of inhabitants in the area is partly explained by the political situation in the neighbouring countries: Since 1994, refugees, especially from Burundi have crossed the border to Kigoma, and in June 2005, there were still ca. 400,000 refugees in the region.[2] In addition, refugees from earlier influxes have already settled in the community.

The Ha area, especially that of Kibondo, is fairly isolated because of poor road conditions. When the political situation in the neighbouring countries was stable, Kigoma used to be an important city on the trade route between Tanzania and Burundi as well as to the Democratic Republic of Congo (Zaire in those days) across Lake Tanganyika, and it was also an important transit port for passengers. Nowadays, Kigoma is one of the least developed regions in Tanzania. The region is best connected to the coast by railway, which, however does not run everyday, and is not very reliable. There are also regular flights to Kigoma.

Ha is an Interlacustrine Bantu language, closely related to Rundi of Burundi and Nyarwanda of Rwanda. These three languages, together with Vinza, Hangaza and Shubi, are classified as group D60 Guthrie (1971), Ha being D66.

---

[1] Census figures from the official website of the government of Tanzania (www.tanzania. go.tz/census/, cited 25th October 2005).

[2] Figures by ReliefWeb (Affected populations in the Great Lakes Region – as at 15th June 2005, www.reliefweb.int, cited 25th October 2005).

In the Ha area, there are several distinct dialects. These dialects show some variation in vocabulary and pronunciation, but are mutually intelligible. Further surveys are needed in order to establish the exact relationships between these dialects as well as with the other languages in the same group.

Ha is a typical Bantu language with 16 proper noun classes and a complex verbal morphology. It is highly agglutinative. The tonal system is analysed as a pitch-accent system. The orthography proposal discussed in this paper is based on the first modern phonological and grammatical description of the language (Harjula 2004).

TOWARDS A LITERARY LANGUAGE

The orthography used in the Ha translations made by missionaries working in the Kigoma area at the beginning of the 20th century followed that of Rundi. These translations include some liturgical literature, such as the Catechism and hymn-books, even portions of the Bible. These translations are old and not in use nor publicly available, according to my knowledge. The Bible translation seems to have gone missing when the missionaries retired and went back to their home countries. Nowadays, the language of the churches is predominantly Swahili.

Since Swahili and English are the languages of education in Tanzania, Ha is seldom written, and the Ha orthography has not been standardised. In 1997, a book of Ha conversations was published by Mwegerano *et al* (*Tugaanire Mugiha*). It introduces conversations and texts of everyday life in Ha, as well as proverbs, riddles and some history of the Ha, accompanied by Swahili translations and vocabulary. The orthography used in this book is modified from Swahili, but it is not consistent.

The linguistic choice for a writing system would naturally be an orthography which marks all the phonemic differences and nothing else, and where a certain grapheme always represents the same phoneme and a certain phoneme is always represented by the same grapheme. A proposal for the Ha orthography is proposed in my PhD dissertation (Harjula 2004). However, in practice there are many other considerations in designing an orthography.

Firstly, the orthography has to be accepted by the speakers of the language. There is no point in designing orthography, however neat it would linguistically be, if it is not accepted and used by the language community. Secondly, the new standardised orthography should be accepted not only by some speakers, churches, etc., but by the whole community, or at least the majority of the community, including the different dialects. This requires negotiations between the representatives of different dialects and sociolects, and possibly also reflects the choice of a central or widely accepted form of the language as a basis for the

literary language. With this in mind, the Kasulu area dialect of Ha is considered to be the 'genuine' version of the language. It is also geographically central and well understood and accepted by speakers of the other dialects. In case of differences between the dialects, the Kasulu form was generally chosen.

Practicality is another aspect in establishing an orthography. There should not be unnecessarily complicated multigraphs nor too many diacritics, nor any other elements that are not needed by speakers of the language. The orthography should be technically feasible, i.e. the combinations of graphemes and diacritics are to be found, if possible, in standard font sets or typewriters. Also, the practical choice is often a writing convention that utilises the same solutions as the other languages in the area, e.g. the alphabet or the tone marking system. Moreover, any possible guidelines from the national or regional language policy makers have to be followed. In Tanzania, the guidelines for new orthographies are set by The Languages of Tanzania project (LOT), University of Dar es Salaam.

Finally, the orthography designed has to be properly tested. This includes monitoring both the readability and writability of the orthography, as well as its acceptability by the community. If problem areas show up, new decisions are needed, and new solutions are to be tested as many times as needed before standardisation.

HA ORTHOGRAPHY DECISIONS

Now that the basis for a standardised orthography, a description of the language, has been made, it is time to take steps to ensure that the Ha writing system is officially recognised. United Bible Societies (UBS) has started a Bible translation project in the Ha area. The three translators elected represent the main dialectal areas, i.e. Kigoma, Kasulu and Kibondo. For the translation to function, the orthography, grammar and vocabulary must be standardised.

In June 2004, UBS arranged an orthography workshop in Kasulu. This workshop was attended by leaders of the churches in the Ha area and other influential people selected by the community, with representation from all the main dialects. The aim of the workshop was to carefully consider both the linguistic proposal as well as the existing writing conventions, and produce an orthography proposal. This proposal, presented in detail in the following, is currently being tested by the translation team.

The graphemes

The proposed graphemes are presented in *Table 1*. The most prolonged discussions were needed on the use of graphemes <r> and <l>, the grapheme <b> versus <bh>, and some palatalised sounds and sound changes.

p		t		k
b		d	j	g
	f	s	sh	h
	v	z		
pf		ts	ch	
m		n	ny	
w		r (and l)	y	
i	e	a	o	u

Table 1: Graphemes of the proposed Ha orthography

The single graphemes are identical to the phonemic symbols (see Harjula 2004), except that /j/ is represented by <y>, and /ɟ/ by <j>. Since [ɾ] and [l] are in free variation, or one or the other is preferred in different dialects, only one symbol, <r>, was chosen for the orthography, to make it easier to maintain a standard way of writing. However, it is expected that both <r> and <l> will be used in practice, at least before the standard way of writing is properly established and recognised.

In some existing writing conventions, [b] and [β] are written with <b> and <bh>, respectively. However, these two sounds are in complementary distribution ([b] occurs after [m] and with labialisation, [β] elsewhere) and could, thus, be represented by a single grapheme <b>. The argument for using <bh> was, in addition to the fact that some speakers had already got used to it, the need to inform the reader about the different pronunciation of the two sounds. Since the differentiation is not needed by native speakers, and for outsiders it will be easy to show the rule of complementary distribution, the committee decided that only the symbol <b> should be used, since the use of <bh> for [β] would be an un-necessary complication.

The digraphs <pf>, <ts>, and <ch> of the orthography represent the affricates /p͡f/, /t͡s/, and /t͡ʃ/, respectively. Some of the participants also wanted to introduce the digraph <ph>, but it was not included in the proposal, since /f/ is more economically represented with <f>.

In Rundi, <c> is used instead of <ch>, but in this matter the committee opted for the same convention as in Swahili and as in the LOT proposal. Both /ɲ/ and /nʲ/ are written with <ny>, and /ʃ/ is written with <sh>.

	Orthographic	Phonemic	Gloss
p	ipeeni	/ipéːni/	'pen'
t	umutaama	/umutaːma/	'elderly man'
k	ibikooko	/ibikookó/	'animal'
b	urubere	/uɾubéɾe/	'millet'
d	daata	/daatá/	'father'
j	ikijiiji	/ikiɟíːɟi/	'village'
g	ibigoori	/ibigóːɾi/	'maize'
f	ifuma	/ifuma/	'foam'
s	isaake	/isáːke/	'hare'
sh	ishuugwi	/iʃúːgʷi/	'flower'
h	ibiharagi	/ibihaɾagi/	'beans'
v	ivu	/ivú/	'soil'
z	izuuru	/izúːɾu/	'nose'
pf	ukupfa	/ukúp͡fa/	'to die'
ts	umutsi	/umút͡si/	'tendon'
ch	ichumu	/it͡ʃúmu/	'spear'
m	akamaro	/akamaɾo/	'benefit'
n	umunaazi	/umunaːzi/	'coconut palm'
ny	nyoko	/ɲoko/	'your mother'
w	iwaari	/iwáːɾi/	'rice'
r	umuraazi	/umuɾáːzi/	'guest'
y	ibuye	/ibuje/	'stone'

Table 2: Examples of the graphemes and their phonemic equivalents

The labialisation is written as <w> after the labialised consonant (1), palatalisation as <y> after the palatalised consonant (2), unless the secondary articulation affects the base sound. Labialised /r/ is written with <gw> (3), palatalised /r/ with <dy> (4), palatalised /k/ with <ch> (5), and palatalised /b/ with <vy> (6), following the pronunciation. In Rundi, the labialised and palatalised /r/ are written with <rw> and <ry>, respectively, and the labialised /b/ with <bg>. Because Ha pronunciation approaches the Rundi pronunciation in the border areas, the Rundi way of writing labialisation and palatalisation were discussed in the workshop. These options were dismissed because the digraphs <gw>, <dy> and <bw> represent central Ha pronunciation better.

(1)	umwe	/umʷé/	'one'
(2)	ukufyaatura	/ukufʲaːtuɾa/	'to fire'
(3)	ugweeso	/uɾʷeːso/	'pot'
(4)	dyaanje	/rʲáːɲe/	'my'
(5)	ichaara	/ikʲáːɾa/	'nail'
(6)	ivyaara	/ibʲáːɾa/	'nails'

Prenasalisation is written before the prenasalised consonant as the nasal corresponding to the place of articulation of the consonant (7)–(8), with the exception that <n> is written before the palatals and the velars in addition to being written before the alveolars (9). Prenasalised /ɾ/ is written as <nd>, according to the actual pronunciation of the combination (10).

(7)	imbeba	/iᵐbeba/	'rat'
(8)	intoke	/iⁿtoke/	'fingers'
(9)	igitunguru	/ikitúᵑguɾu/	'onion'
(10)	ndamenya	/ⁿɾaméɲa/	'I know'

Dahl's Law affects the concord prefixes with voiceless plosives /t/ and /k/, voicing them when the stem-initial consonant is voiceless. This voicing is marked in the orthography.

(11)	iki+ha	igiha
	CP7+'Ha'	'Ha language'
(12)	tu+fise	dufise
	SP1pp+'have'	'we have'
(13)	uku+suuguur+a	ugusuuguura
	CP15+'harvest'+FV	'to harvest'

Tone

Since the pitch-accent is contrastive in Ha (according to analysis of the tonal system in Harjula 2004), speakers sometimes have to read a clause twice or three times to get the tonal melody right when accents are not marked. This particularly happens with grammatical tones, as lexical minimal pairs are rare. Thus, linguistically the marking of the tone would be desirable, and technically marking the surface H with an accent (´) would not be a problem. LOT recommends that all contrastive elements, including tone, should be marked.

But, in practice, the marking of tone is often found to complicate reading and especially writing too much, since the speakers of the language have not got used

to seeing the tone marked. The weight of the lexical tone being so light in Ha, and the grammatical tone being usually predictable by the environment, the marking of tone was not introduced to the proposed Ha orthography. After all, the orthography has to be accepted, above all, by speakers of the language.

Vowel lengthening

The LOT recommendation is to mark phonemic long vowels by doubling the vowel grapheme. In Ha, this would mean that only the long vowels of the noun and verb stems which are not caused by a consonant-approximant or nasal-plosive cluster would be marked. This way of marking would probably work for reading, but, in writing, people should be able to distinguish the stems from the inflectional elements in order to determine whether to write the vowels as short or long.

In addition to the contrastiveness of long vowels in stems, vowel lengthening is the only surface manifestation of the first person singular object prefix *n-* which is deleted before a nasal-initial verb stem. This should be marked in the orthography by a double vowel if any long vowels are marked.

(14)	wó+tu+meny+*+a	wotumenya
	RPro1+OP1pp+'know'+M2+FV	'who knows us'
(15)	wó+n+meny+*+a	woomenya
	RPro1+OP1ps+'know'+M2+FV	'who knows me'

The index forms *na**, *nka**, and *nga** and the connexive *-a** lengthen the augment of the following word when there is an accent on the following mora (16)–(17).

(16)	na*+iN+nyáabu	niinyaabu
	A1+CP9/10+'cat'	'with a cat'
(17)	na*+iN+taama	nintaama
	AI+CP9/10+'sheep'	'with a sheep'

There are also instances where automatic phonetic vowel lengthening is overruled by other circumstances. The stem-initial vowel of the verb stems is not lengthened when it is preceded by the first person singular object prefix *n-* (18). All the word-initial and word-final vowels are short (19), with the exception of the independent relative forms where the prefix consists of a vowel only (20).

(18)	uku+erek*+a	ukweereka
	CP15+'show'+FV	'to show'
	a+n+erek*+a	anyereka
	SP1+OP1ps+'show'+FV	'he shows me'

(19)   iN+bwá   imbwa
        CP9/10+'dog'   'dog'
(20)   uu*+ki+kor*+a   uukikora
        SPRel1+Pers+'work'+FV   'the one who is still working'

When taking into consideration the problem of deciding on the contrastive long vowels, it seems reasonable to either write all long vowels or to write all vowels short. The linguistic choice between these two options would obviously be to mark all long vowels. The language committee had a lengthy discussion about the need of marking the long vowels at all, and at the end they decided to go for marking the long vowels. But it was noted already during the workshop that some practice is needed before writing the long vowels will be consistent. Practical tests will show whether marking the long vowels will just cause trouble or will be helpful.

Vowel elisions

Since vowels are often deleted at morpheme and word boundaries in normal speech, it had to be decided which vowels are to be marked and which left out in writing the Ha language. In Nyarwanda and Rundi, the elided vowels are marked with an apostrophe ('), but additional marking of morphophonemic processes is not recommended for the languages of Tanzania by the LOT project. However, the language committee strongly insisted that the elided vowels should not be written anywhere, but marked with an apostrophe. Their argument for this decision was that the elided vowels are never pronounced in normal speech. The marking of the elided vowels by apostrophes might cause trouble in consistency and in the automatic processing of the language, and should thus be carefully tested in practice.

The augment is deleted if it is not noun phrase initial, i.e. if it is preceded, e.g., by a locative prefix, the negative existential copula *ntaa*, or a noun phrase initial pronoun. These elements are to be written separately from the noun they modify (21)–(23). Since the secondary locative prefixes cannot occur independently, they are written in conjunction with the nouns (24). The same applies to the relative pronouns (25) and the dependent substitutives (26).

(21)   ntaa umu+ntu   ntaa muntu
        NegEx CP1+'person'   'there is no one'
(22)   uyu umu+ntu   uyu muntu
        DemP1 CP1+'person'   'this person'

(23)	uwu+ndi* umu+ntu	uwundi muntu
	CP1+'other' CP1+'person'	'another person'
(24)	mu+uru+gereero	murugereero
	Loc18+CP11+'market'	'at the market'
(25)	hó+uku+ichar+a	hokwichara
	RPro16+CP15+'sit'+FV	'where to sit'
(26)	ntaa+chó	ntaacho
	NegEx+Pro7	'there is nothing'

Numerals are usually used without the augment, so the augment does not need to be written with numerals. The augment is also deleted from the second part of a compound noun, so these should be written together without the augment.

Final vowel deletions of the verbs and the verbal copulas are to be marked in the orthography by an apostrophe (27). The connexive -*a** and the index forms *na**, *nka**, and *nga** all lose their vowel and affect the augment (28)–(29). These are written in conjunction with the following element even if it does not have an augment (30)–(31).

(27)	ní+umu+áana	n'umwaana
	Cop+CP1+'child'	'it is a child'
(28)	na*+umu+áana	n'umwaana
	AI+CP1+'child'	'with the child'
(29)	i+a*+umu+áana	y'umwaana
	DP9+Con+CP1+'child'	'of the child'
(30)	umu+twé u+a*+u+eéne	umutwe waweene
	CP3+'head' DP3+Con+DP3+Pro	'its head'
(31)	na*+nje	nanje
	AI+Pro1ps	'with me'

More problematic is the marking of final vowel elisions of verbs and nouns. With the marking of the elisions, the word forms are written in different ways, depending on the context. This might cause problems for consistency and thus make it more of a challenge to the automatic processing of the language, e.g. spell checking. The linguistic choice would be to write the elided final vowels as such, since fluent readers would be able to drop the relevant vowels in reading. But the language committee made a decision to mark the elided vowels by an apostrophe.

(32)	mpa iyo inkoni	mp' iyo nkoni
	OP1+'give'+FV DemP9 CP9+'stick'	'give me that stick'

(33)   aragira inka   aragir' inka

         SP1+Foc+'herd'+FV CP9+'cow'   'he herds cows'

Proper names are written with initial capital letters. Otherwise, proper names are written as common nouns.

(34)   i+a*+i+Kibondo   y'iKibondo

         DP9+Con+CP23+PropN   'of Kibondo'

FROM NOW ON

The orthography proposed by the workshop is a compromise between linguistic ideals, practical issues and acceptability. Some issues, such as the marking of vowel elisions, were linguistically argued against, but acceptability ruled over the linguistic, and perhaps, the practical solution. In other instances, such as the unnecessary distinction between <b> and <bh>, the representatives were ready to adjust their personal preferences and customs.

As for now, the orthography proposal is being tested by the translation team. However, the work is really time consuming, and it will still take years before Ha speakers have an officially standardised orthography for their language. Even then, it remains up to them to follow the standard or not.

Abbreviations

AI	associative index
Con	connexive
Cop	copula
CP4	noun class prefix of class 4
DemP4	demonstrative pronoun of class 4
DP4	determiner prefix of class 4
Foc	verbal focus marker
FV	final vowel
loc	location
M2	tense accent assigned to the second syllable of the macrostem
NegEx	negative existential copula
OP4	object prefix of class 4
Pers	persistive marker
Pro7	noun class 7 substitutive pronoun
PropN	proper name
RPro	relative pronoun
SP4	subject prefix of class 4
SPRel8	independent relative prefix of class 8

BIBLIOGRAPHY

GORDON, Raymond G. Jr. (ed.) 2005. *Ethnologue: Languages of the World.* 15th ed. Dallas: SIL International.

GUTHRIE, Malcolm 1967–71. *Comparative Bantu,* I–IV. Farnborough: Gregg International.

HARJULA, Lotta 1999. Kiha Phonology. M.A. thesis, Institute for Asian and African Studies, University of Helsinki.

------   2004. *The Ha Language of Tanzania: Grammar, Texts and Vocabulary.* Köln: Rüdiger Köppe Verlag.

MWEGERANO, Anderson M. J., Nestory D. KIPFUMU & Pamphili CHUBWA 1997. *Tugaanire Mugiha. Tuzungumze katika Kiha.* Salo, Finland: Salon Painokeskus Reiote OY.