Bantu Orthography Manual

Leila Schroeder

SIL International

Part 1

Bantu Orthography Manual

For linguistic and literacy fieldworkers

Abstract

The Bantu Orthography Manual is a resource for developing writing systems among the Bantu subgroup of Niger-Congo languages. It offers a strategy for orthography development, combined with a list of resources for Bantu linguistic information and the condensed advice of a coterie of respected Bantu linguistic experts. It offers readability and write-ability considerations whenever applicable. The Manual has a target audience: linguists gathering information for orthography development.

Procedures for a "participatory approach" to phonological analysis are described in one chapter. A series of charts help the linguist document and organize the phonological and morphological information gathered. The Linguistic Features chapter lists common linguistic characteristics of Bantu languages, describes the attendant orthographic challenges, and offers suggested solutions, along with the pedagogical rationale for each.

Contents

Part 2: A Participatory Approach to Orthography Development: Procedural Guide

Appendices

References 278

Foreword

The Bantu Orthography Manual is not a computer tool but a resource to aid in orthography decisions. It offers a suggested strategy for orthography development with a workshop approach, combined with a list of resources for Bantu linguistic information and the condensed advice of a coterie of respected Bantu linguistic experts: Rod Casali, Myles Leitch, Oliver Stegen, Constance Kutsch-Lojenga, Ken Olsen, Mike Cahill, Karen Van Otterloo, and Bill Gardner.* It offers readability and write-ability considerations whenever applicable, and is written with an informal style to facilitate easy access to information.

Part 1, Bantu Orthography Manual, has a target audience: the Ordinary Working Linguist (OWL). [1] As such, it does not offer the degree of linguistic detail one can find in such extensive, deep and informative collections of Bantu linguistic information as Nurse and Philippson (2003), though it refers the reader to many such resources.

Part 2, Participatory Approach to Orthography Development provides step-by-step instructions for carrying out a participatory approach to orthography development, a concept originated by Constance Kutsch-Lojenga. Its style is less formal and very practical. The process involves workshops with groups of participants possibly representing several related Bantu languages.

While the Manual has an academic focus, the Participatory Procedural Guide is less formal and very practical. It attempts to guide an OWL through a participatory approach to orthography development. It hyperlinks to the Bantu Orthography Manual, which should be used often as a reference, as well as a series of eight Literacy Lessons, plus Proto-Bantu morphology charts and a large Orthography Workbook Form. All are designed for use during a workshop.

The Linguistic Features chapter of the Manual provides suggestions to help the linguist and mother-tongue speakers choose symbols to represent the linguistic features of their languages. It gives an overview of many significant phonological and morphological processes as they interact in Bantu languages. Examples of the orthographic challenges are given and advice is offered as to how these phenomena might best be reflected in a writing system. [2] This chapter can also be used as a complement to the Bantu Phonology Tool, a computer tool for phonological analysis of Bantu languages.

The principles and suggestions given try to incorporate pedagogical, perceptual, and sociolinguistic considerations into the decision-making process for graphemic choices. In this way, it is hoped that the writing system for a given Bantu language will become the possession, the useful tool, of the mother-tongue speaker.

---

*Special thanks go to these contributors: Rod Casali, Constance Kutsch-Lojenga, Bill Gardner, Myles Leitch, Karen Van Otterloo, Oliver Stegen, Keith Snider, and Helen Eaton.

[1] The phenomena described focus on East and Central Africa Bantu languages, with the hope of publishing a future edition that reflects a wider area. Swahili is often referred to as a typical language of wider communication. However, the procedure can be adapted to other parts of the Bantu-speaking world, where other languages function as the LWC.

[2] Native speaker perception and preferences are embedded in our participatory approach to phonological analysis, which precedes the graphemic decisions, and with orthography testing in local communities (addressed in other chapters).

Chapter 1 - Defining Orthography

Definition of orthography

Consists of two parts:

1. Symbols (in our case, alphabets)
2. Writing rules

   - Alphabets require linguistic input from the Bantu dictionary and phonology tools
   - Writing rules need these, plus input from Bantu grammar and discourse tools
   - Both alphabets and writing rules need input from psycho-linguistic, sociolinguistic, educational and political realities in the context

The purpose

- Empowerment of mother-tongue speakers to read and write their languages as a result of their own informed orthography decisions
- Advancement of academic knowledge and professional skills, both our own and those of our partners

The goal

Develop an orthography for a Bantu language group, in partnership with speakers of that language (laying a linguistic and educational foundation for the application of that orthography to reading and writing)

Foundational assumptions for orthography development

- Community ownership/involvement in every stage of the orthography development process, not only to develop orthographies which are accepted by local communities, but to encourage local understanding of the rationale underlying orthography decisions
- The centrality of linguistic analysis to orthography development, inclusive of phonology, morphophonemics, grammar and discourse (since these aspects of a spoken language overlap and influence one another on the surface, they will seem to compete with one another for prominence in a writing system)
- Mother-tongue speakers' perception should play a significant role in orthography decisions. That perception can be developed and enriched for those who take part in the orthography development process
- Orthography-in-use as goal, and also as means for constant feedback and evaluation
- Questions of readability and write-ability will be considered throughout the development process

- Revisions will be ongoing, and will necessarily reflect the political, educational and social context of the writing system

Elements of the development process

- The Orthography Development tool is more a concept or a process than a finely tuned mechanism; i.e., it will not require a computer program

- Templates for the products to be produced while applying the tool/concepts could be made (i.e., for a Word book, an Alphabet Book, and/or a Writers' Guide or transition primer).

- Workshops to produce these and plans for testing followed by application in communities are the actual tool (or the format/vehicle for implementation)

- Interaction with Ordinary Working Linguists, or OWLs (people who are now developing and testing their linguistic analyses and applications), to see what challenges they have encountered in their work and what info (and wisdom) they've gained as a result.

A suggested sequence for such an approach is given in Chapter 5, with activities listed under each objective. The attendance and involvement of local language speakers in each activity described is assumed; ordinary working linguists and data entry personnel will also be involved in most of the activities listed.

Purpose of this manual

This orthography manual is not a complete guide for phonological analysis, but is designed for applied linguistics, specifically orthography development. It is intended as an aid toward development of trial orthographies for Bantu languages, using a workshop approach. It assumes the reader has some familiarity with the grammatical and phonological characteristics of Bantu languages, and is therefore brief in its descriptions of some Bantu linguistic phenomena. It contains references, as well as footnotes, and a list of suggested background reading materials. The last appendix in Part 2 (Appendix L: Orthography Workbook Form) is a worksheet to be filled in during a series of two workshops.

Chapter 2 - Representing Linguistic Features

Decisions to be made at a Writers' Workshop

Introduction

This chapter provides suggestions to help the linguist and mother-tongue speakers choose symbols to represent the linguistic features of their languages. It gives an overview of many significant processes that occur in Bantu languages and how these phenomena might best be reflected in a writing system. We begin with an acknowledgement that linguistic features are not the only factors in orthography development: "On the surface, orthography selection and development are linguistic issues; but in practice they are loaded with imperatives arising from a number of sources" (Eira 1998:171).

Orthographies are complex visual representations of language and thought which are designed to facilitate literate communication. Because of the intertwining of phonology, morphology, and several cognitive processes in the decoding of print, sociolinguistic, and pedagogical considerations for orthography decisions will be presented in this chapter, though it is organized around linguistic features of Bantu languages. Just representing sounds with graphemes is not enough. Language is more than sound, and reading and writing are more than recognition and transcription of sounds alone. A good orthography enables readers to quickly recognize meaning, and its spelling rules are as simple as possible, to aid the writer.

Solid linguistic analysis, while not the only factor in orthography decisions,[1] is essential to the development of a good orthography. It under-girds any writing system which accurately represents the speech and perception of the mother-tongue speaker. Toward that same end, the following principles and suggestions try to incorporate pedagogical, perceptual, and sociolinguistic considerations into the decision-making process for graphemic choices. In this way, it is hoped that the writing system for a given Bantu language will truly become the possession, the useful tool, of the mother-tongue speaker.

William Smalley's five criteria for orthography decisions (1964:34) are still widely used today, though he had transfer from mother-tongue to the language of wider communication (LWC) in mind. In Africa today, our situation is often the reverse, with mother-tongue literacy coming later. His criteria with my annotations are:

1. Maximum motivation for the learner (Smalley refers to issues of ethnic and national identity)

2. Maximum representation of speech (phonemic writing, with some exceptions where morphology affects phonology, and vice versa)

3. Maximum ease of learning (simplicity of rules)

4. Maximum transfer [to/from] literacy in a LWC, and

5. Maximum ease of reproduction (for publication purposes)

---

[1]"A better use of linguistics is as a source of insights about orthography options, and as a tool to probe the orthographic insights of native speakers" (Bird 2000:29).

One further point could be added:

6. Maximum recognition/transmission of meaning.

Smalley alludes to preservation of morphemes in his second point. Bantu languages, though, are particularly capable of packing multiple morphemes within one word. Phonological processes alter the surface structure of those adjoining morphemes significantly. Since consistency of morpheme shape facilitates spelling, reading fluency, and comprehension of written text, a Bantu writing system's ability to help the reader recognize morphemes merits focused attention.

Suggested Foundational Resources

This orthography manual is not a guide for phonological analysis, but for applied linguistics, specifically orthography development. It is necessarily specific in its focus; it is not a substitute for general knowledge of orthography principles. I advocate familiarity with Bird (1998), Gardner (2001, 2005), Kutsch-Lojenga (1993), Smalley (1964), Snider (1998, 2005), and Ssemakula (2005). These sources will give you some generalizations about the phonological processes and morphological structures of Bantu languages or explain basic orthography principles not detailed in this manual. A course on orthography principles is also offered in the summers at the University of North Dakota.

Principles Concerning Representation of Segments

A word about shallow versus deep orthographies is necessary here. There is a continuum between deep, meaning-based writing systems in which the primary units of meaning represent morphemes (such as Mandarin), and more shallow, alphabetic writing systems which are based on phonemes, such as Italian, Serbo-Croatian, and Luganda. Deep orthographies represent the morphology of the language more than they do the phonology, while shallow orthographies aim to closely represent the phonology of the language. Even within alphabetic writing systems a variety of depth exists, between the more morpheme-preserving orthographies such as English and French and the more phoneme-based orthographies of Swahili and Lugungu.

An ideal orthography matches the language it represents. "Shallow orthographies are characteristic of languages in which morphemic relatives have consistent pronunciations" (Mattingly 1992:150). At the opposite end of the spectrum are languages which require a deep orthography. They have morphemes which are subject to a lot of phonological variation and/or many homophones which must be distinguished in writing.

There is a tension between the strengths of each kind of orthography. The shallowest orthography gives the new reader/writer lots of control. With recognition of a handful of symbols he/she can recognize and spell anything they can speak or say, based upon a one-to-one correspondence between sound and symbol. But no writing system is actually capable of completely representing speech. Both readers and writers are assisted by writing according to word-level phonology, rather than phrase-level phonology. Moving towards a deep orthography allows homophones to be distinguished while words whose pronunciation varies in context (the dog vs. the apple) can be given a fixed representation. This allows the experienced reader to access meaning rapidly. The issue of morphology and its effect upon the phonology of a language (morphophonological alternation) will arise very frequently in the process of orthography development for a Bantu language.

Orthographies should usually be developed with the mature reader in mind. Therefore, preserve morphemes when possible, especially if a surface-level representation seems to be

in equal competition with a morpheme for graphemic representation. Think beyond the individual segment. Think of a whole word. Lexical changes which take place within a word are usually written. (For more explanation, see Part 1, Appendix D.)[2] Lugungu, for example, has a rule that changes the phoneme /r/ into the phoneme /d/ when a morpheme ending in /n/ is prefixed. That morphophonemic change, /kuruga/>/nduga/ when the prefix /n/ is added, takes place within the boundaries of a word and should be written.

When it's possible (sociolinguistically) to avoid digraphs, do so—but you must consider the transferability of your symbols to the LWC (language of wider communication) in your area. If the LWC is Swahili, your spelling options are rather narrow.

Spell the way people speak in slow speech, rather than rapid conversation. Elision of segments dropped in rapid speech is not recommended for several reasons:

1. If a phoneme or a morpheme is written fully, people can recognize meaning quickly. For example, the Bantu associative marker is /Ca/ or /CCa/. The initial consonant of this word changes depending upon the class of the noun it modifies, leaving only the word-final /a/ to give it consistency of appearance. In rapid speech, this one vowel is often elided and replaced, if the noun it precedes is vowel-initial. For example, in Ikizu (E402J), the associative marker has an underlying <a> as its vowel, but because it is pronounced together with the following word, in rapid speech the <a> coalesces with a vowel-initial augment and elides. The associative, a word in its own right, is written with the vowel that would ordinarily be the augment of the following noun. Visually, then, recognition of the associative marker is completely lost to the reader whenever it precedes a vowel-initial noun.

2. If a phoneme is written fully, ambiguity is kept at a minimum. For example, if a certain vowel is elided in rapid speech, if it is not represented in writing, the reader may be unsure which vowel is missing.

3. The apostrophe, generally the symbol of choice for indicating elision, is used in the standard Swahili alphabet as a part of the trigraph <ng'>. When the LWC includes this trigraph, mother-tongue reading fluency is slowed and accuracy of spelling may be hindered if the apostrophe is further used to indicate elision or phonological attachment of clitics. Apostrophes have several uses in many Bantu orthographies, sometimes indicating a glottal stop (in Cameroon), the attachment of clitics to larger words, the deletion of a vowel across word boundaries, or as part of the consonant digraph <ng'>. This multi-purpose usage of <'> slows fluency of reading and hinders accuracy of spelling. Avoid such multiple uses of a symbol.

At times (rarely) for sociolinguistic reasons, an orthography may have to represent a phonetic sound [ ], rather than just representing the contrastive phonemes / /, of a language. For example, a language group may insist upon writing a distinction between [l] and [ɾ] for purposes of reading transfer to a LWC, even though they may not be contrastive phonemes for their language group. The symbols for sounds are enclosed in [ ] and the graphemes in < >.

---

[2]This principle takes on great importance when spelling and word break decisions are made, for associative markers in particular (also see section on word breaks).

The best grapheme options are listed first; less desirable ones follow them. The format for each grapheme suggestion will be like this one:

[ɑ:]   Options: < **aa** >

List of graphemes for the segmental aspects of a Bantu writing system

Vowels

1. Vowel Systems

2. Five-vowel systems: < **i   e   a   o   u** >

3. Seven-vowel systems: < **i   e   ɛ   a   ɔ   o   u** > or < **i   ɨ   e   a   o   ʉ   u** >

4. Vowel Length: < **VV** >

5. Semivowels: < **y** > and < **w** >

6. Vowel elision: see Vowel elision

Consonants

1. Liquids

    [l]   Options: < **l** > or < **r** >

    [ɾ]   Options: < **r** > or < **l** >

2. Fricatives

    [ð]   Options: < **d** > or < **dh** >

    [Ө]   Options: < **th** >

    [ɣ]   Options: < **g** > or < **gh** >

    [ß]   Options: < **b** > or < **bh** > or < **v** >

    This list is not exhaustive

3. Nasals

    Velar Nasal

    [ŋ]   Options: < **ŋ** > (or < **ng** > or < **ng'** > or < **ñ** >

    Palatal nasal [ɲ] (see "Labialization and palatalization" below)

    Homorganic nasalization

4. Affricates

    [tʃ]   Options: < **c** > (or < **ch** > or < **tch** > or < **tsh** >)

    [ts]   Options: < **ts** > or < **tch** > (less desirable Francophone)

[dz]   Options: <**dz**>

[dʒ]   Options: <**j**> (Swahili influence) or <**dj**> (French influence)

[ʒ]   Options: <**zh**> or <**jh**>

5. Implosives

[ɓ]   Options: <**bb**> or <**bh**> or <**b**> (or <**'b**> or <**b'**>)

[ɗ]   Options: <**dd**> or <**dh**> or <**d**> (or <**'d**> or <**d'**>)

6. Double articulated stops (Labialized velars)

[k͡p]   Options: <**kp**> or rarely <**pk**>

[g͡b]   Options: <**gb**> or rarely <**bg**>

7. Geminate consonants

8. Prenasalization - Prenasalized labiodental fricatives

[ᵐf]   Options: <**mf**> or, less optimally, <**nf**>

[ᵐv]   Options: <**mv**> or, less optimally, <**nv**>

9. Prenasalization - Prenasalized palatals

[ⁿt͡ʃ]   Options: <**nc**>

[ⁿd͡ʒ]   Options: <**nj**>

10. Prenasalization: Prenasalized velars

[ⁿk]   Options: <**nk**> or <**ñk**>

[ⁿg]   Options: <**ng**> or <**ñg**>

11. Prenasalization: Prenasalized double articulated stops

[ⁿ ᵐk͡p] Options: <**nkp**>

[ⁿ ᵐg͡b] Options: <**ngb**>

12. Labialization and palatalization - Labialized consonants

[tʷ]   Options: <**tw**> or <**tu**>

13. Labialization and palatalization – Palatalized consonants

[tʲ]   Options: <**ti**> or <**ty**>

14. Labialization and palatalization - Palatal nasal

[ɲ]   Options: <**ny**> or <**ɲ**> or <**ñ**>

15. Palatal nasal contrasting with palatalized alveolar nasal [nʲ] and palatalized palatal nasal [ɲʲ]

[ɲ]   Options: <**ny**> or <**ɲ**> or <**ñ**>

[nʲ]   Options: <**ni**> or <**ny**>

[ɲʲ]   Options: <**nyi**> or <**nyy**>

Labialized and palatalized consonants contrasting with CSV sequences

[kʷe]   Options: <**kwe**> or <**kue**>

[kuwe]Options: <**kuwe**> or <**kue**> or <**kuue**>

[pʲa]   Options: <**pya**> or <**pia**>

[pija]   Options: <**pya**> or <**piya**>, <**pja**> or <**pija**>

16. Velarization: see Velarization

Vowels

Since there are only five vowel letters in the Roman alphabet, we are left with the option of altering these currently existing letters to multiply the grapheme options.

Bantu orthographies most commonly have added ~ or ˆ over a vowel to distinguish it from a similar grapheme, or put a line directly through it, as in <ʉ>. Underlining and, rarely, the placing of a dot under a vowel have also been used. These latter two options make letter/sound recognition slower, but with nine-vowel systems the need for contrastive representation requires ingenuity. The ability to type the characters is important for literacy, but several Unicode-compatible options are now available.

Bantu phonology is highly sensitive to morphological considerations. Underlying vowel distribution within specific morphological slots is highly restricted, both in seven-vowel and five-vowel languages. Not only are vowels limited as to their distribution within a word, but there are rules which restrict their occurrence in sequence, especially within stems (Hyman 2003:46).[3] Typically, there are more contrastive vowels stem-initially than in the rest of a Bantu word.

Vowel systems

Five-vowel systems

The standard five Roman alphabet letters are used for these systems, representing the vowels closest to those of the LWC in the perception of mother-tongue speakers.

		Front	Central	Back
	High	i/ɪ		u
Swahili's five-vowel system is shown below (Maddieson:16)	**Mid**	e/ɛ		o/ɔ
	Low		a	

Written form	Phoneme	Swahili	English
i	i	**tisini**	*ninety*
e	ɛ	**-chelewa**	*be late*
a	a	**kwanza**	*first*
o	ɔ	**kiboko**	*hippo*
u	ʊ	**ugali**	*porridge*

Seven-vowel systems

Two kinds of seven-vowel systems are typical, with contrast for advanced tongue root (for simplicity's sake, we are considering *+/-ATR* distinction as equivalent to *tense/lax*).

---

[3]For an example, see Figure 1, Tharaka Vowel Coalescence.

Seven-vowel system (lower type)

		Front	Central	Back
High	**+ATR**	ɪ		ʊ
Mid	**+ATR**	e		o
Mid	**-ATR**	ɛ		ɔ
Low	**-ATR**		ɑ	

Seven-vowel system (upper type)

		Front	Central	Back
High	**+ATR**	ɪ		u
High	**-ATR**	I		ʊ
Mid	**-ATR**	ɛ		ɔ
Low	**-ATR**		ɑ	

Rangi (F33), Nyakyusa (M31), Nande (DJ42), and Vwanji (G66) are examples of the upper type seven vowel system. Oliver Stegen (pc) provided the following examples from Rangi.

Written form	Phoneme	Rangi example	English
i	ɪ	**irino**	*mushroom*
ɨ	I	**ɨmbwɨ**	*spider*
e	ɛ	**kɨnene**	*abdomen*
a	ɑ	**isalala**	*drizzle*
o	ɔ	**ilolo**	*fallow field*
ʉ	ʊ	**ʉnʉ**	*mosquito*
u	u	**nguru**	*strength*

A word about the distribution of vowels in seven-vowel systems may be helpful.

Affixes often contain a smaller vowel inventory, most typically five vowels. They may exhibit phonetic variants with the five phonemes, though, if they harmonize with the vowels in the roots. The vowels in the augment and prefix actually can help prepare the reader to recognize the root.

Word-finally, even roots of words may have a smaller vowel inventory. There is often even devoicing of these word-final vowels.

Nine-vowel systems

Bila (D32), like Budu (D332) and some others in northeastern DRC, has a nine-vowel system with ATR-based vowel harmony (Kutsch-Lojenga 2003:450-474). A nine-vowel system will have to distinctively represent the vowels /ɛ/ and /ɔ/.

	Front	**Central**	**Back**
+ATR	i		u
-ATR	ɪ		ʊ
+ATR	e		o
-ATR	ɛ	ɑ	ɔ

Vowel Harmony

Vowel harmony is a common phenomenon in which a vowel or vowels harmonize in some quality with other vowels in a word. It can be seen as the spreading of a phonetic feature (back, high, round, advanced tongue root) within a word. The spread is usually leftward from the root, or leftward from a derivational extension. The vowels of one group tend to occur with each other in words, to the exclusion of vowels of the other group. The phoneme /a/, however, is usually a member of both groups in Bantu languages, so it is not affected.

Vowel height harmony is very common, and usually affects only the vowels of derivational suffixes such as the causative, applicative, reversive/separative, neuter, and combinations of these.

ATR harmony, on the other hand,.occurs only in Bantu languages with seven or more vowels. The vowels are divided into two mutually exclusive groups, one known as +ATR vowels, and the other as -ATR. As stated earlier, the vowel /a/occurs with both groups. It often blocks the spreading of the harmony in a given language.

In a language with ATR harmony, all roots (noun roots, verb roots and others) will contain only vowels of one group or the other, but not both. The "dominant" value of ATR (usually +ATR in narrow Bantu) is the one that causes the other value to change to be like it. In a system where +ATR is dominant, a -ATR vowel can change to become +ATR, but a +ATR vowel will never change to become -ATR.

In most Bantu languages, the +ATR feature spreads leftward from a triggering vowel. This may be a root vowel affecting a prefix as in (2), or a suffix vowel causing a change in the vowels to its left. This means that even the vowels in the root can be affected.

Vowel height harmony is shown in (1) with altered vowels in bold, and ATR (advanced tongue root) harmony as reported for many Bantu languages such as Rangi (F30), in (2) from Stegen (pc), and Malila (3) from Helen Eaton (pc) and Constance Kutsch-Lojenga. The Rangi example shows that the –ATR nominal prefix is changed to +ATR when it occurs with a +ATR root.

(1) Swahili:

-imba	'sing'	+ applicative >	**-imbia**
-kata	'cut'	+ applicative >	**-katia**
-vuta	'pull'	+ applicative >	**-vutia**
-leta	'bring'	+ applicative >	**-letea**
-soma	'read'	+ applicative >	**-somea**

(2) Rangi:

mʊ-	work	mʊrɪmɔ
	snake sp.	mʊlalɔ
	mouth	mʊlɔmɔ
	God	mʊlʊ^ŋgʊ
mu-	slope	mugiritʰɔ
	whistle	muluri

Kutsch-Lojenga says of Malila (M20), "The [+high] initial vowel of the -ile suffix is [+ATR] and changes all other vowels of the verb word to its left into [+ATR], however many /a/-vowels are found in between. This means changes in the verb-root vowel from /ɪ/ to /i/ and from /ʊ/ to /u/ as well as the the allophonic realisations [e] and [o] of underlying /ɛ/ and /ɔ/." The allophonic differences between e /ɛ and o/ɔ are not written (note that the /a/ is not affected by the harmony process).

(3) Malila

kʉmʉpɨmba   'to carry him'

amupimbile   'he has carried him'

kʉmʉbhʉʉzya   'to tell him'

amubhuziizye   'he has told him'

Mother-tongue speakers are aware of the changes from one vowel phoneme to another, in these examples. This means the changes should be represented in the orthography. Some changes are only allophonic, and these are not written. In both ATR languages exemplified above, when a +ATR vowel follows underlying –ATR mid vowels /ɛ/ and /ɔ/, these mid vowels are pronounced as [e] and [o]. These distinctions, only discerned by the phonetician, are of course not written.

Write word-internal changes which the native speaker is aware of, even if this means a given morpheme, such as a noun class prefix or even a verbal root, may have two written forms .[4] That would mean that word-internal phonemic (not phonetic) changes due to vowel harmony rules should be written. This rule will make spelling intuitive for the writer.

The reader is already accustomed to the morphophonologically conditioned alternations in speech, so will learn to access meaning despite the resulting altered appearance of some morphemes. The reader is often actually assisted by this surface-level representation of vowel harmony, because it usually moves leftward from the root. The reader, beginning at the left end of a word and moving to the right, finds a lot of predictability of the form the root will take, based upon the first vowel he/she sees. Indicating the vowel harmony

---

[4]A rare result of vowel harmony which alters a vowel in the root is ambiguity between two roots, but the context usually clears it up for the reader and the writer is helped immeasureably by writing what he/she actually hears.

changes also means that the reader does not have to read the entire word before he can pronounce its beginning correctly.

The following section offers suggestions for orthographic representation of the vowels of a given language.

1. In choosing graphemes for each of the vowels, decision-makers must first consider which of them are most closely associated with those of the LWC. These will probably be given commonly used vowel symbols. The remainder then will probably require diacritics to distinguish them. Possible diacritic modifications to a vowel symbol are: <←>, <û>, and <ū>. In writing systems which represent tone above any syllable, the first option, <←>, is preferable to facilitate recognition of both the vowel symbol and its tone. The Tharaka use of a circumflex (<û>, <î>) above two of its vowels would be inappropriate if tone were written over any of its vowels (it is not).

2. If the orthographies of surrounding languages use certain diacritics for vowels, this may affect your choice of symbols, especially if many people have learned to read using that system.[5]

3. For consistency, always use diacritics with the *same* + or − ATR attributes, because the contrast will be a natural one for the mother-tongue speaker. For example, *do not* use a diacritic on a [-ATR] front vowel, while also using it on a [+ATR] back vowel.

4. One suggestion, *all other things being equal* (i.e., never violate rule 3), is to add the diacritics to the least frequently occurring vowels. This facilitates speed for typists and writers, since the symbol won't have to be written as often. It may also facilitate speed of symbol recognition for readers.

The suggested vowel graphemes are shown below, and they avoid diacritics above the letter.

V Options

5V orthography: < i e a o u >

7V orthography: either < i e ɛ a ɔ o u > or < i ɨ e a o ʉ u >

9V orthography: < i ɨ e ɛ a ɔ o ʉ u >

Vowel length

Some Bantu languages have phonemic vowel length, while others do not.

Hyman (2003:42) lists five [potential] sources of vowel length in Bantu:

---

[5]This assumes that the number of vowels in these nearby languages is equal to the number in the language in question. Underdifferentiation of vowels for a language should be avoided at all costs!

1. Phonemic length

2. Concatenation (either across morpheme boundaries or through consonant elision)

3. Gliding plus compensatory lengthening

4. Compensatory lengthening preceding a moraic nasal plus consonant

5. Penultimate vowel lengthening

This chapter addresses point 1 under "Phonemically long vowels." It addresses points 2, 3, 4, and 5 briefly, under "Conditioned long vowels."

Rod Casali identifies two more sources of vowel length.

1. elision of one of two adjacent nonidentical vowels, with compensatory lengthening of the remaining vowel (for an example, see the Tharaka vowel coalescence chart, figure 1

2. phonetic lengthening of underlying short vowels that bear surface contour tone

In regard to this last point, it should also be noted that a long vowel (either phonemically long or one in which length phonologically conditioned or attributable to morphological concatenation) may be longer, or at least perceived as longer, when it is realized with a contour tone (falling or rising) rather than when it has a level tone.

Phonemically long vowels

Some Bantu languages have phonemic vowel length, while others do not.

Proto-Bantu has been reconstructed as having seven vowels /į,[6] i, e, a, o, u, ų/ plus phonemic length (Hyman 2003:42). Many, but by no means all, Bantu languages have retained phonemic length in their vowel systems. It is easy to determine whether a language has phonemically long vowels by sorting bisyllabic (CVCV) noun and verb stems using the participatory approach to language analysis. If there is phonemic vowel length, these stems will fall into two sets: CVCV and CVVCV. Noun stems containing consonant-glide sequences (Cy or Cw) or nasal-consonant sequences (NC), e.g., mb, nd, ng, are not used in this initial sorting for phonemic length. They will be used in the sorting for conditioned vowel length. If the sorting of the stems with glides and prenasalized consonants also results in two sets, there is no orthographically significant conditioning for vowel lengthening, and all long vowels must be marked as such.

If a language has phonemically long vowels, they must be marked in the orthography as long. The orthographic representation of phonemic vowel length is doubly important in languages where the marking of phonemic tone is done sparingly, because the fewer phonemic cues the reader has, the harder it will be for the reader to determine the meaning of a given word.

Most Bantu orthographies denote long vowels by doubling the vowel symbol. The use of two identical vowel symbols seems to match Bantu speakers' perception of vowel length: <aa>, <ee>, <ii>, <oo>, <uu>.

---

[6]The vowels į and ų constitute extra-high vowels. The subscript diacritic is a convention peculiar to Proto-Bantu vowels.

People do not generally object to the writing of double vowels once they realize that vowel length is an important contrast in their language. The use of double vowels will be unfamiliar even to most literate people, however, and at first may make it harder for people to spell their language correctly.[7] Thus, spelling rules must be taught and practiced once the orthography is introduced. With time, readers and writers will realize the usefulness of long vowels in distinguishing words whose contrast would be otherwise ignored by the writing system.

Phonologically conditioned long vowels (in languages with phonemic length contrast)

In many of the Bantu languages with contrastive (phonemic) vowel length, one also finds noncontrastive (phonologically conditioned) vowel length.

Conditioned vowel length is usually found in these environments:

- before a prenasalized consonant
- following a labialized or palatalized consonant (Cy or Cw) or a semivowel (y or w)[8]

If a language has phonologically conditioned length, as a great number of Bantu languages do,[9] speakers will significantly lengthen any vowel in the two environments mentioned above. Speakers will not be able to distinguish a short vowel from a long vowel in these environments. They will all sound the same length to them. Thus, there will be no vowel length contrast possible in this environment.[10] Whether the vowels in these conditioning environments should be written as short or long in the orthography depends on a number of factors and is best determined by testing the perceptions and preferences of both readers and writers.

Depending on the language, the actual phonetic duration of such conditioned long vowels may be nearly the same as a phonemically long vowel, or it may be something in between. Ganda (JE15) and Sukuma (F21) both have phonemic length and conditioned length. However, the conditioned long vowels in Ganda are much closer in duration to underlying long vowels in Ganda than they are in Sukuma. Sukuma lengthened vowels are almost exactly intermediate between underlying long and short vowels (Maddieson 2003:37).

In some languages, perhaps like Sukuma, speakers may feel that the vowels in the length-conditioning environments group with short vowels, rather than long. If this is the case, the obvious choice is to use a single vowel to write the phonologically conditioned long vowels, since it matches the speakers' perception, and they will automatically apply the correct degree of phonetic lengthening to vowels which occur in the conditioning environments.

An example of a language where conditioned vowels are not quite as long as phonemically long vowels is Tharaka. However, whenever a vowel precedes a prenasalized stop, the vowel still sounds long to a mother-tongue speaker. Thus, actual phonetic length of these

---

[7]During literacy lectures at the first orthography workshop, it is useful to give examples from Dutch, English, or other prestigious languages which do manifest double vowels in their writing systems (e.g., English <**look**>, <**feed**>).

[8]Sometimes a Cy or Cw will condition following length, but a simple glide (w or y) will not. All environments should be checked before the linguist decides where vowel length is contrastive and where it is not.

[9]According to Hyman (2003:52), "In most Bantu languages there is no vowel length opposition before an NC complex."

[10]A possible exception would be the result of morphemes rather than phonological processes.

vowels is not necessarily an indicator of how they are perceived by speakers. However, despite their perception that the conditioned length vowels are long, because this length is not contrastive, the Tharaka orthography does not write the conditioned length with double vowels. Thus, readers must be taught (requiring a lesson on prenasalized stops) to write such vowels as short.

In languages where speakers feel that the vowels in these environments are clearly long, writing this length, even though it is not contrastive, may make spelling easier for them. Mother-tongue speakers may want to write all the vowels which sound long to them as long. This is one of the options for representing conditioned long (CL) vowels.

But because speakers will automatically produce the conditioned length in the conditioning environment, they can often best be written with single vowels, even when they are perceived as long. If the orthography choice is made to represent CL vowels with a single letter, writers must be taught to write such vowels as short, despite their perception (requiring a literacy lesson on prenasalized stops).

This is the choice taken by the official Ganda (EJ13) orthography. A note in the orthography statement or worksheet, with examples, will be sufficient to explain the simple environments that condition the automatic lengthening.

Example: In Ganda, following semivowels and preceding prenasalized stops, vowels are always written short (http://www.buganda.com/language.htm).

> [tuunda] *'sell'* is written < **tunda** >
>
> [kitaanda] *'bed'* is written < **kitanda** >
>
> [kiimpi] *'it is short'* is written < **kimpi** >.

There are pros and cons to each option, and these vary from language to language. They should be discussed by representatives of the language community and choices should be tested by a literacy specialist and/or a linguist before a final decision is made.

The disadvantages to writing all long-sounding vowels as long:

- Using double vowels to represent noncontrastive length makes words longer.
- If both contrastive and noncontrastive length are indicated orthographically, truly contrastive length will not stand out to the reader like it does when double vowels are used *only* to indicate contrastive length.
- In some cases, truly contrastive length which results from morpheme concatenation will be masked by the use of double vowels for the representation of conditioned vowel length. The following examples can be found under "Other vowel length issues".

The disadvantages to writing all vowels with conditioned length as short:

- In languages where tone markings are used frequently, the use of a single vowel to indicate a vowel with conditioned length can complicate the representation of falling or rising tone. Such contour tones can be indicated by a sequence of a high tone mark and a low tone mark for falling tone (or the opposite order for rising tone) when they occur on a long vowel indicated by two letters. This allows all tones to be marked using only two symbols. But if a contour tone occurs on a conditioned long vowel which is represented by a single vowel, it is necessary to

introduce a circumflex and/or a breve or caron to indicate the contour tone(s).

Discussion of these issues and a good degree of testing the tendencies and perceptions of both readers and writers will be necessary to determine whether or not conditioned length should be written with two vowels or one.

Languages with no true conditioned length

In some languages in which there is phonemic vowel length, this length is always contrastive,[11] even in the environments where length is conditioned in many other languages. In such a language, either a long vowel or a short(er) vowel may occur in the conditioned environments, and the long vowels still contrast with short vowels no matter what the phonological environment is. In such a case, long vowels must always be written as long, and short ones as short. This is apparently the case in Matuumbi (Odden 2003:532) as well as in Rangi (F33), where there is contrast preceding a prenasalized consonant in the following two noun stems:

Example: Rangi: **mʉsinga**   *'child'*   vs.   **mʉriinga**   *'beehive'*

We will now look in some detail at various spelling challenges for languages with conditioned lengthening (all highly typical of Bantu languages) and discuss the orthographic considerations relevant to each.

Vowel length from morpheme concatenation

Sometimes, vowel length (or a series of two consecutive vowels in separate syllables) results at a morpheme boundary, from the juncture of a vowel-final morpheme and a vowel-initial one.

Example: Bungu (F25), which seems to have no phonemic length in stems (Stegen:4), exhibits a contrast in verb tenses that needs to be represented by the use of a double vowel. This conjoining of two vowels often only happens where the verb stem is vowel-initial, as in the example here, where the stem is -anda. (See a similar example from Ndali (M301) later in this section.)

Example: Bungu   <**naanda**> *'I am beginning'*   vs.   <**nanda**> *'I have begun'*

As Stegen states, long vowels can result from cross-morphemic vowel sequences, at times involving vowel assimilation in case of nonidentical vowels. This can occur through direct concatenation as in the following.

Example: Rangi   **éékire** *'he has shut [implied: the door]'*   (from á + ék + ire)

Or it may involve consonant elision as in the following example.

Example: Rangi **ámwíímbíire** *'he has sung for him'*   (from á + mú + ímb + ir + ire)

In most cases, such long vowels will be written with double letter, as they can be semantically distinct from the same segmental sequence with short vowel, as demonstrated in the example below in contrast to the example above.

---

[11]This should not be confused with contrastive length that occurs in these environments only in a few selected words, occasioned by morpheme concatenation or found in one or two specific grammatical morphemes.

Example: Rangi: ámwiimbíré *'he should sing for him'* (from á + mú + ímb + ir + é)

However, since Rangi may have contrastive long vowels in all phonological environments, all their long vowels are written with a double vowel.

Helen Eaton (pc) writes concerning languages in the Mbeya area of Tanzania, "We are finding that for those languages where lengthened vowels in conditioned environment are perceived by MT speakers as just as long as underlying long vowels, the morphologically lengthened vowels also sound just as long, and no longer." Take Malila (M24), for example.

1   < **kʉlô̱nga** > *'to say'*

2   < **nâ̱nda** >   *'I have begun'*   (from na + anda)

3   < **né̱ega** >   *'I have taken'*   (from na + ega)

4   < **ábhi̱nza** > *'cl.2-good'*   (from abha + inza)

5   < **ábhi̱ilu** > *'cl.2-black'*   (from abha + ilu)

 Eaton says, "The underlined vowels are all considered the same length by MT speakers. If we taught that the morphologically lengthened vowels before NC (2 and 4) needed to be written long, that would match up with forms like 3 and 5, but differ from forms like 1, in which the vowel is perceived as just as long."

As long as there is not any meaning contrast lost by writing all conditioned length vowels with a single letter, it is simplest if the rule can be maintained without exception.

Cases which necessitate exceptions to rules

Real languages are complex, and even in the extremely regular Bantu languages, there will be cases which necessitate an exception to your spelling rules. Do not question your rules just because you find an exception or two. If you find extensive and systematic sets of exceptions, consider revising the rules.

Individual exceptions to the rule of writing conditioned length with a short vowel may need to be made in cases of verbs with vowel-initial stems. This does not mean that the "write them short" rule is incorrect or invalid. The exceptions will involve a meaningful contrast, and this fact may make teaching them fairly concrete. Testing will be needed to determine the best option for differentiating otherwise ambiguous words or morphemes.

Example of a lexical contrast needing an exception to a rule:

In Tharaka (E54), vowels are always lengthened about 1½ beats before prenasalized stops. Their orthography follows the practice of writing a conditioned long vowel with a single letter. This means that they have a spelling rule which says not to write a double vowel preceding a nasal plus consonant. However, the following words are contrastive in meaning (quite possibly the second word is actually a clitic plus another morpheme):< **îndî** > *'but'* and < **îîndî** > *'now'*. Since they have different meanings, they need to be written differently, as vowel length (and tone) in the second word indicates a

very distinct lexical item. Thus, they need an exception to their rule, saying that in the word <**îîndî**> '*now*' a double vowel must be written, even though it precedes a prenasalized consonant.

Note that if instead the Tharaka writing system simply required that any vowel that sounds long must be written with a double vowel, both of these words would be written <**îîndî**>. The contrast in meaning between the two words would be lost.

Second example of a lexical contrast requiring an exception to a rule:

In Ndali (M301), strict adherence to the rule of writing vowels short in a conditioned length environment results in neutralized contrasts, such as that between <**akanda**>'*he has stamped*' from /a-kanda/ and <**akanda**> from /a-ka-anda/ '*and he started*'. In such a case it will probably be best to teach readers and writers to reflect the morphemes which are otherwise involved in the loss of contrast, writing /a-ka-anda/ as <**akaanda**> in order to maintain the semantic difference between the two forms. This can be taught in a lesson focusing on vowel-initial verb stems, which often cause exceptions to length rules, due to the concatenation of a vowel-final tense or person morpheme and the vowel-initial verb stem.

More than two vowels in a row

In a case where concatenation results in three vowels in a row, writing them as three consecutive vowel letters may not be accepted. The following Tharaka (E54) verb in the distant past exemplifies such a situation. The person marker is ba-, the past marker is a- and the root is -ak-, resulting in three <**a**>'s in a row.

Example: Tharaka: /baaakire/ '*they built (long ago)*' (from ba + a + ak + ire)

The Tharaka example /baaakire/ is written orthographically as<**baakire**>, because they decided to disallow triple sequences of identical vowels in the orthography. Kithinji et al. give the following explanation.

> In some cases it is possible that three identical vowels occur in sequence…
> In such cases all three vowels will not be written but only two. For example, ba + a + akire, 'they built' is written baakire (not baaakire).
> This means that in the surface form this will not be distinguished from
> the recent past, ba + akire, [also] spelt <**baakire**>. The two forms are
> pronounced with different tone patterns, and the context will determine
> in most cases which form is intended. (Kithinji et al. 1998:3)

Especially if it can be shown that there is a difference in the syllable pattern of such a word, writers may be convinced that writing all three vowels is actually a good option.

Vowel length attributable to the loss of intervocalic consonants

In a few Bantu languages, phonemic length has not been retained from Proto-Bantu, but contrastive length has developed from the loss of intervocalic consonants or morpheme concatenation. In such cases, there will not be any phonologically conditioned length. Among these is Tonga (M64) (Maddieson 2003:38).

Vowel length only in certain morphemes

It is also possible that a contrastive long vowel may be retained in certain tense prefixes even in a language which otherwise has no phonemic vowel length (Nurse 2003:100). This is especially true of length due to morpheme concatenation or the loss of an intervocalic consonant, but can also be due to certain retained or regained vocabulary items. Botne (2003:425) notes that in Lega (D25) long vowels occur most commonly in ideophones (which generally tend toward onomatopoeia and intonational length) and rarely in other words—he lists only three other words that have long vowels. Thus, even in a language with no systematic phonemic vowel length, there can be the need for an occasional double vowel. This would call for the singling out of such morphemes as teaching points in an orthography guide and any literacy pedagogy.

Sequences of nasal plus consonant which do not condition preceding vowel length

In some languages, there are certain specific vocabulary items which contain sequences of nasal plus consonant that do not condition length in the preceding vowel. Often these are not reflexes of Proto-Bantu NC sequences, but rather reflect the historical loss of a vowel, usually <u>, from a <mu-> prefix. In other cases, they involve a prenasalized continuant, such as nj, nz, rather than a prenasalized obstruent. The presence of such non-length-conditioning sequences does not rule out the use of a rule that a vowel be written as short preceding a prenasalized consonant. Testing should be done, especially with readers, in such cases.

Predictable shortening of an underlyingly long vowel

There are also many cases in Bantu languages in which underlying long vowels undergo predictable shortening in some contexts:

A number of languages allow a long vowel only in the penultimate or antepenultimate syllables, and shorten any long vowel which would occur before that. Some languages, like Safwa (M25) and Kifuliiru (DJ63), count morae rather than syllables and allow no more than two morae maximum to follow a long vowel. This sort of shortening usually applies both to conditioned vowel length and to phonemically long vowels.

This may present the need for a spelling rule by which any phonemic length in a verb stem is preserved in stems which would otherwise lose their contrastive length when several suffixes are added. It is not an issue in nouns and other words which do not have significant suffixation.

The following example (Van Otterloo, pc) shows that from the literacy point of view, it is sometimes preferable to write underlying length rather than reflecting phonological shortening in verbs with contrastive length in the stem.

Example: Fuliiru (DJ63) has two verb stems with lexical contrast for vowel length,<**kubiika**> *'to put'* and <**kubika** > *'to crow'*. <**kubiika**> will lose the vowel length in its radical, phonetically, whenever there are three or more morae following in its suffixes, e.g., with the long causative /–iis/ plus the final vowel. With the three following morae, the stem vowel shortens. [-biik-] is shortened to [-bik-] in [kubikiisa] *'to cause to put'*. If the shortened stem vowel is written with a single letter, the contrast will be neutralized orthographically between the two verbs.

The underlying length in the stem should be preserved if testing verifies that readers can learn to identify the verb root, and hence meaning of the whole word, in this way. The verb radical may be concrete enough that writers would be able to deal with maintaining non-phonetic length there (Van Otterloo). In this case, then, the verb with causative [kubikiisa] would be written <**kubiikiisa**>.

The other implication of the presence of a shortening rule is that in an orthography which represents conditioned vowel length using double vowels, a large percentage of stems would be affected by any length-preservation rule such as that mentioned in the previous paragraph. Instead of involving only stems having phonemically long vowels (usually a relatively small number) it would also theoretically involve any stem in which there is a phonologically conditioned long vowel.

If a spelling rule preserves the morphemic shape of a verb stem such as [-biik-] above, this will either necessitate also preserving length in a verb stem such as [-geend-] (in which the vowel length is merely conditioned, and not contrastive) or else readers and writers will need to be taught to differentiate between phonemic and conditioned length when deciding which length to retain. If the decision is made to retain all length, testing will need to be done to determine whether it is possible for readers and writers to identify the verb root and to determine that in its unsuffixed form it exhibited phonological vowel lengthening. This extrapolation might be difficult enough that it would outweigh the ease factor of a decision to write all vowels that sound long with a double letter, and be a factor in deciding to write conditioned long vowels with a single letter in all instances.

Summary recommendations for writing vowel length

- Languages with phonemic length contrasts write single versus double vowel letters, e.g., <**a**> vs. <**aa**>.
- Long vowels resulting from morpheme concatenation should be written with double letter, at least in cases where meaningful contrast would otherwise be neutralized.
- Conditioned length vowels which are perceived as short should be written with single letter.
- Vowels which occur in an environment which conditions length in the language, but which are perceived as extra long (i.e., there is a tone change across the vowel, or the length results from morpheme concatenation and is perceived as longer than that of monomorphemic conditioned length vowels in the same environment) should generally be written with double letter, subject to testing of readers and writers. Such words will need to be taught as exceptions to the rule of writing a single vowel in a phonological environment which conditions vowel length.

Semivowels

Mother-tongue speakers' perception of a vowel as a consonant in certain contexts will probably determine your choice of symbols for semivowels. Again, however, certain ambiguities can be created when a phonemic (and morphemic) distinction is lost when replacing a certain vowel with the semivowel. For a five-vowel language this is rarely a problem. /i/ is always written as <**y**> before any other vowel besides itself, and /u/ is always written as <**w**> before any vowel besides itself, as shown below.[12]

---

[12]A morpheme is lost, though, when two identical ones adjoin, as can be seen in /mi+i/ <miri> and /u+u/, <unga> in the Duruma example above.

Examples from Duruma (E72d) nouns, showing what happens at the juncture of noun class prefixes and vowel-initial roots:

class 4 pl: mi-

mi	+	iri	→	**<miri>**	*bodies*
mi	+	ezi	→	**<myezi>**	*months*
mi	+	adine	→	**<myadine>**	*trees, sp.*
mi	+	oho	→	**<myoho>**	*fires*
mi	+	uho	→	**<myuho>**	*rivers*

class 14 sing: u-

u	+	ira	→	**<wira>**	*song*
u	+	embe	→	**<wembe>**	*razor blade*
u	+	ari	→	**<wari>**	*food from maizemeal*
u	+	ongo	→	**<wongo>**	*brain*
u	+	ungs	→	**<unga>**	*flour*

In seven-vowel languages, on the other hand, things can be more complicated. See Figure 1 Tharaka Vowel Coalescence, for an example. Both /ʊ/, spelt < **û** >, and /u/, spelt <**u**> in certain contexts can be pronounced or perceived as consonantal. If a <**w**> were written in both instances, an important lexical contrast would be lost! The decision was made to write <**û**> before any vowel besides itself as <**w**>, but to write /u/ before any vowel besides itself as <**u**>.

Surprisingly (since the higher of the two vowels would seem to function more like a palatal), the same loss of distinction takes place with glide formation after either /i/ or /î/ within a Tharaka word. In the case of /i/ and /î/, people chose to write /î/, the lax front unrounded vowel, as <**y**> before any vowel besides itself. The /i/ remains the same in the writing system, before any vowel besides itself. This is common in Bantu languages, including Swahili.[13]

As with the prenasalized stops, Tharaka semivowels usually bring about (conditioned) length—in the vowels immediately *following* them. The following is an example (Kithinji 1999:5).[14]

> Watho, 'law', will sound as if a long aa sound is present. But we know that the root of the word is -atha, to which is added the (abstract) noun class prefix û, which changes to <w> before a. Since the underlying form only has one vowel, only one vowel will be written. This is still pronounced long.

---

[13]Rod Casali (pc) says, "I have a strong suspicion that the [ATR] distinction…is, in many seven-vowel languages, not neutralized phonetically either, but that what has commonly been transcribed as [y] is really a high front vocoid…and that what is transcribed as [w] is really a high round vocoid." Or, there may be a "greater desyllabification with the [-ATR] than the [+ATR] vowels," with the contrast not really neutralized phonetically. If this is true, the Tharaka orthographic choices made perfect sense!

[14]I have altered the formatting of the original article so that the symbols for phonemic, phonetic, and orthographic representations are consistent in this manual.

The same applies to the other semi-vowel<**y**>. When this occurs before a single vowel, that vowel will sound as if it is long. For example, <**yake**>, 'his' is made up of /î + -ake/. It sounds long but there is only one /a/ vowel present in the underlying form.

Here are some other examples.

û+ onokio	=	<**wonokio**>	*salvation*
û + ega	=	<**wega**>	*good*
î+akwa	=	<**yakwa**>	*mine*
î + enu	=	<**yenu**>	*yours (pl.)*

When two identical vowels follow a semi-vowel, then both vowels will be written. This is a long syllable and will sound long with or without the presence of the semi-vowel. For example, '*you built*' (distant past) is <**waakire**>. Here the different parts making up the word are /û + a + aka + ire/. Here we see that one <**a**> vowel is the past tense marker, and the other <**a**> is the first letter of the word '*build*'. Other examples are the following.

û + a+ anjîîria	=	<**waanjîîria**>	*he started*
û + a + athaga	=	<**waathaga**>	*he reigns*
î + a + amba	=	<**yaamba**>	*it spread*
î + a + akire	=	<**yaakire**>	*it applied itself*

Also note that two vowels that join together to form one long vowel will also both be written after a semivowel, like the following.

| wa + ûgu | = | <**woogu**> | *they said* |
| ya + ûgu | = | <**yoogu**> | *it said* |

So basically, the Tharaka spelling rule is: where there is a SV (semivowel) syllable which only has one vowel in the underlying structure, then only one vowel will be written, but where there are two vowels in the underlying form, then two vowels will be written. In pronunciation (the surface form) it is possible that both will sound the same. A trial period, and eventually testing, may be necessary to determine whether people can grasp series of rules such as this. The Tharaka adults found their vowel length rules to be very difficult to master at first. The children, though, who received instruction in Tharaka grammar as part of their spelling exercises, may be mastering it.

This is another example of phonological rules impacting the surface representation of morphemes. In this instance, in order to preserve an important semantic contrast, the morpheme, or underlying form, was represented in the orthography.

If a noun root starts with a vowel, for example Kwaya, E251(J) /eki-aanga/ '*skull*', pronounced [ekyaanga], there may be an underlyingly long vowel root initially (*regardless* of the prenasalized stop which follows it). Mother-tongue speakers are acutely aware of its presence.

What shows us the underlying length?

The common occurrence of long vowels in this position in the language, even where there is no possibility of semivowel insertion.

The existence of two morae, shown by a tone change across a long, or double, vowel.

Recommendation: Write both of the double vowels which follow the vowel in the prefix, if there is an underlyingly long (contrastive with short V-initial roots) vowel in this position. The written word will exhibit three vowels in a row, so people may find it helpful to attach it to the noun it precedes, with a hyphen. The prefix is usually a clitic. Insertion of a hyphen breaks up the 3-vowel sequence visually, and sets apart what amounts to two words. This choice may benefit the writer, giving more simplicity and consistency in spelling rules for prefixes. It preserves the integrity of true vowel length, which is underlying, not conditioned. The writer doesn't have to remember the complicated rule, "Never write a long vowel after y or w, or before prenasalized consonants, *except* after a prefix."

Example: <**aka-uuna**> or <**ekya-anga**>

Never insert a semivowel between two vowels unless speaker perception calls for it. Otherwise you make recognition of morphemes which are indicated via these symbols, such as the causative or the passive, harder for the reader, and you make words longer.

Vowel elision (also a word break issue)

If at all possible, avoid using an apostrophe and deleting a vowel in a writing system, even if the vowel is rarely spoken. Usually, the issue of insertion of an apostrophe arises when a short word comes before a V-initial word. Since in rapid speech the vowel at the end of the short word is dropped out, mother-tongue speakers want this vowel dropped in the writing system. Doing this creates several problems for the reader.

There may be another /t-/ particle in the language which can occupy the same position in a sentence. In Tharaka, removing the /i/ from the representation of the negative form of the verb 'to be', /ti/, may leave its interpretation ambiguous. Maybe the letter elided was /a/ instead, giving <**t'**> a completely different meaning.[15]

Elision makes word recognition harder for the reader in another way: it visually unites two words with very disparate meaning and function, making it harder to recognize the noun or adjective to which the smaller word is now attached (while normally such words stand on their own).

Adding an apostrophe after an <**n**> also creates unnecessary challenges for rapid recognition of the commonest symbol for the velar nasal, <**ng'**>. Since apostrophes are already in use with this digraph, and a <**g**> is the only intervening contrastive symbol, it is advisable to avoid the apostrophe in marking elision.

---

[15]The Tharaka particles /ta/ and /ti/ are examples. The first of these particles means like, as. The second is the negated form of /i/, *is not*. In the orthography it was decided that the first would always be written as <ta>, despite its elided pronunciation before a V-initial word. On the other hand, /ti/ is written in full before consonants, and regularly as < t' > before vowels. So, though the appearance of the morpheme undergoes change, at least there is no ambiguity between ti and ta because ta is always fully written out. Writing the full form of /ta/ was a deliberate measure to maintain its distinction from /ti/ (Kithinji 1999:5).

Last, if there are already tone symbols or vowel diacritics written above the writing line, adding apostrophes will add to the complexity of eye movement and word/grapheme recognition.

Note: Where elision is a fast-speech phenomenon, there will be considerable pressure upon the linguist to write "the way people speak." It may help local people who want to write "the way we speak," to see the need to at least preserve the morpheme for *one* of the possible elisions. People may agree, for example, to always write out <**a**>, but to elide the <**i**> in /ti/, as in the Tharaka example above. Don't expect new literates to have the finesse we experienced literates have, in distinguishing the correct word from the context! When there's a question (and an option), spell out morphemes as they're pronounced in slow speech.

On the other hand, if the underlying vowel is lost completely, even in slow speech, do not force the issue or you'll make the language very hard for people to write. Choices can still be made based upon what people know about their language. If a form has been fully elided and become semantically one, such as <**mwanwakwa**> *'my child'*, it can be written as such, with no apostrophe needed. In this case, two words have merged semantically, forming a compound, a new morpheme which is an amalgamation of the two.

Preserve word image, especially the smaller grammatical words. As Scott Satre (pc) says, "Americans don't always pronounce the <**h**> in 'his/her' in the middle of a sentence in English, but we always or almost always write it, depending on the 'register' of the communication." Mother-tongue speakers should not have to rely upon sounding out everything they read orally in order to get meaning from a text. Even though Americans also say "jago (e.g., "tʌ thʌ stor)?" in rapid speech, the separate words <**did you go**> are written as such.

Hyphenation: When you have a series of words which are semantically closely connected but not merged, and it's lengthy for the reader, consider hyphenation, as in *sister-in-law*.

Consonants

Liquids

[l] and [ɾ] or [ʈ]

Because of borrowed words from other languages, some Bantu languages have a contrast between /l/ and /r/. If this is the case for the language in question, both should be written. Many Bantu languages have a single phoneme with allophonic variation between [l] and [ɾ]. It is optimal to write only one, especially if most of the mother-tongue speakers are not already reading a language which distinguishes the two in writing. If there is one liquid, <l> is usually chosen (especially if it is already in use with a LWC).

[l]	Options:	<l> or <r>

[R]	Options:	<r> or <l>

Example: In Zinza E23(J), /ɾ/ does not occur as a phoneme contrasting with /l/, but <r> is written, in accord with Roman Catholic tradition, for proper names.

Fricatives

In many cases, voiced Bantu fricatives become stops following a nasal so one grapheme represents the allophonic variants, unless, in rare cases, people are acutely aware of the distinction. This may happen in places where a LWC such as Kiswahili is very influential.

Plosive b [b] contrasting with a fricative b [ß]

Gungu, E101(J), is one language in which plosive b and fricative b are contrastive. The fricative is written <**b**>, while the plosive is written by doubling it<**bb**>. This decision was made because the fricative b is more common and corresponds to the plosive /b/ in the orthographies of surrounding languages (Moe 2002:2).

[b]   Options: <**bb**> or <**b**>

[ß]   Options: <**b**> or <**bh**> or <**v**>

Test people for their preferences. Some factors to consider are 1) frequency of occurrence; 2) transferability issues (plosive b is written with a single <**b**> in the European languages and Swahili); and 3) mother-tongue speaker perception, e.g., in many Bantu languages, /b/ is a weak consonant. Pre-nasalization strengthens the voiced bilabial: the fricative [ß] and the stop [**b**] are allophones in this case.[16]

Spelling options become narrower if the language also has an *implosive* bilabial stop. Mother-tongue speaker perception of the three sounds (plus their use in the language of wider communication) will probably determine the choice between the symbols as listed.

Nasals

Palatal nasal [ɲ] (see "Labialization and palatalization")

Velar nasal [ŋ]

Many Bantu languages have a phonemic velar nasal [N] which in Swahili and many other languages is written with the trigraph <**ng'**>. In order to avoid the long, visually complex trigraph, and because the symbol exists on typewriters, some have chosen the symbol <**ñ**>. Another problem with the trigraph <**ng'**> is created by the apostrophe. Its use in a digraph may inhibit fluent reading because apostrophes will also be used to represent elision, commonly after an <**n**>! The simple IPA symbol <**ŋ**> is an excellent representation for this simple phoneme. However, an important consideration in selecting a grapheme for the velar nasal is transferability to the nearest LWC (Alas for languages spoken in Tanzania!).

[ŋ]   Options: <**ŋ**> (or <**ng**> or <**ng'**> or <**ñ**>[17]

---

[16] People's perception of sounds which are allophones for their language may be so acute that the distinction must be written, in rare cases. Overdifferentiation in the writing system may be necessary, if spellers are consistently confused when they write their mother tongue. This may occur when the influence of the LWC is very strong among a language group. An example of this is the Kwaya language, in which velar nasal fricative hardens to a stop after a nasal. For this language, speakers insisted that its fricative variant be written distinctly, with a <gh>." All the related language groups were able to use one symbol for what is actually one phoneme, /g/.

[17] Shona languages of Mozambique (S10) use <n'>, which has the advantage of being a digraph instead of a trigraph, but the disadvantage of looking like a syllabic n (which Shona does not have (Bill Gardner, pc).

Examples:

In Ngbaka (an Ubangian language) /ŋ/ is written as <**ng**> and /ŋg/ as <**ngg**>.

In West African orthographies, a simple<**ŋ**> represents this phoneme. The problem with reproducibility of this symbol may be diminishing with increased computer use.

In Swahili <**ng'**> represents the velar nasal, as in <**ng'ombe**> *'cow'*.

Note: The biggest challenges for writing the velar nasal arise when it is contrastive with a prenasalized velar stop. If this is your challenge, see Prenasalization.

Homorganic nasalization

Some Swahili noun class prefixes (i.e., for classes 9 and 10) consist of homorganic pre-nasalization (nonsyllabic) preceding voiced consonants. /N-boga/ *'vegetable'* is written <**mboga**>, /N-dege/ *'bird'* is written<**ndege**>, /N-jaa/ *'hunger'* is written <**njaa**> and /N-goma/ *'drum'* is written <**ngoma**>, for example. Since the homorganic change is word-internal, the change is written. Voiceless obstruents may undergo the same process.[18]

Exception: In Swahili, the velar nasal /ŋ/, a phoneme in the language, is written by a trigraph <**ng'**>. However, when a velar consonant is prenasalized, the simple letter <**n**> is used, to keep words visually shorter and simpler. For example, /Ngoma/ *'drum'* is written <**ngoma**> rather than <**ng'goma**>.

Usually, prenasalization for Bantu languages is written with <**m**> for bilabial se-quences <**mb**>,< **mp**> and <**n**> precedes the stop for alveolar <**nd**> <**nt**>, palatal <**nj**>,<**nch**> and velar <**ng**>, <**nk**> sequences. For more on pre-nasalization issues, see Prenasalization.

Examples from Lingala, C36d (Ken Olson, pc):

<**mboka** >	[mbókà]	*'village'*
<**ndako** >	[ndákò]	*'house'*
<**ngomba** >	[ŋgómbà]	*'hill'*

Voiceless nasals

Bantu languages with voiceless nasals represent these in their orthography, e.g., in Gogo <**mh**, **nh**, **nyh**, **ng'h**>. An alternative is using a strike-through of the voiceless nasals, as <**m̶**> if the orthography is already full of digraphs, or if the consonant *sequence* /Nh/ also occurs in the language.

[N̥] Options (voiceless nasal):   <**mh**> or <**m̶**>

Syllabic nasals (usually heterorganic nasalization[19])

Syllabic nasals are often word-initial in Bantu, usually predicative morphemes or realiza-tions of the prefix **mu-**. They usually arise from elision of the vowel in a prefix or clitic.

---

[18]Another assimilative process is described by Meinhof's Law, or the Ganda Law. The post-nasal consonant is nasalized and assimilates to the place of articulation of the nasal, thus creating a geminate nasal. See the glossary for an example.

[19]In nasal-consonant sequences which are heterorganic (having two different points of articulation), the

The simplest way to represent syllabic nasals, and to keep them visually distinct from prenasalized sounds, e.g., /^{m}b/ or / ^{n}d/, and from other nonsyllabic nasals, is actually to write the elided vowel in the case of the prefix, but this is not often done. [20]

Example 1: noun classes < > (syllabicity unwritten)

Swahili only has syllabic bilabial nasal: /m̩/, as in /m̩tu/ <**mtu**> '*person*' and /m̩ti/ <**mti**> '*tree*'. The /m̩/ is a noun class prefix attached to a monosyllabic stem. Monosyllabic noun stems make the prefix syllabic.

For Swahili adjectives, classes 1 and 3 are distinguished from classes 9 and 10 by syllabicity vs. homorganicity, e.g., <**mdogo**> /m̩dogo/ '*little [implied: person]*' vs. <**ndogo**> /ⁿdogo/ '*little [implied: object]*'. The syllabicity distinction is not written.[21]

Example 2: Sena (N.40) has minimal pairs for syllabic vs. homorganic nasals, according to Barbara Heins (pc).

<**mbale**>	'*clay plate, pot lid*'	/^{m}bale/	<**m'bale**>	'*brother*'	/m̩bale/
<**nkhondo**>	'*war*'	/ⁿkʰoⁿdo/	<**n'khondo**>	'*path*'	/n̩kʰoⁿdo/

Heins says, "The pronunciation of ph, th, kh, d and b is affected by prenasalization (the <h> indicates aspiration). <d> and <b> are normally implosive and lose their implosivity, as in '*clay plate*' above. In the aspirated stops the consonants are in the process of dropping out as in Shona. <**Nkhabe**> '*no*' is pronounced [N] with a prominent [ʰ], as in '*war*' above."

[N̩]   Options: <N̩> or < > or <**N** + elided vowel> or <**N'**>

Discussion: the apostrophe is probably the most commonly used marker to distinguish syllabic nasals from others, but apostrophes have their problems, being used for several other things, such as elision, possessives and as components in trigraphs. Zero marking may work if minimal pairs, such as those in Swahili, are few.

Affricates

Voiceless palatal affricate [tʃ]

Due to the influence of European languages, the voiceless palatal affricate [tʃ] is often spelled **ch**. Orthographers have sometimes opted for the simpler <**c**>.

---

nasals are always syllabic. Additionally, syllabic nasals with homorganic sequences occur if the nasal is a class 1 or class 3 prefix, e.g., m-bepari '*capitalist*', m-bembe '*seducer*', m-beregezaji '*filibuster*', m-bono '*castor-oil plant*', while in noun class 9, all homorganic NC-sequences are truly prenasalized, e.g., mbinu '*method*', mbolea '*manure*', ndoo '*bucket*,* and ngamia '*camel*'. Despite minimal pairs such as cl.3 m-buni '*coffee plant*' versus cl.9 mbuni '*ostrich*' the contrast is not written (Oliver Stegen, pc).

[20]For clitic grapheme suggestions, see section on Word Boundary Principles.

[21]Karen Van Otterloo (pc): The difference is that classes 1 and 3 have an underlying m- prefix, while 9/10 has an underlying N- (unspecified nasal) prefix. The difference is not audible when the adjective begins with a bilabial consonant: all four classes would be m-bichi ('unripe/wet/raw'). They probably wouldn't go with class 1, but could go with either 3 or 9/10. The difference is that the class 9/10 prefix is a nasal that always assimilates to the point of articulation of the C, while the class 1 or 3 prefix is m-. Example: m is a prefix in m̱bwa '*dog*,* and is simply word-medial in *kwamba* '*that*'.

It might be better to say the contrast is due to word position. An initial NC sequence is usually pronounced with a syllabic nasal, while a word medial NC is pronounced as a nonsyllabic homorganic nasal.

[tʃ]   Options: <**c**> (or <**ch**> or <**tch**> or <**tsh**>)

When a language has prenasalization which must be written (see prenasalized palatals), thus adding to the alphabetic (and digraph) inventory, it's especially helpful to keep the number of graphemes down to a minimum with the <**c**> choice.

Another reason for using <**c**> rather than <**ch**> occurs when there is phonemic aspiration, usually written with an <**h**>. <**ch**> can then represent the aspirated voiceless palatal affricate, while <**c**> represents the nonaspirated voiceless palatal affricate, e.g., in Ndau (S15a). Hence, a simple <**c**> is recommended when possible, sociolinguistically.

Examples from Southern Bantu (Ndau S15):

      <**chuchururu**> *'dry beans'*   /s^hu s^hururu/

      <**ciwi**>   *'island with reeds'*   /tʃiwi/

Alveolar affricates [ts] and [dz]

These segments are not uncommon in Bantu C, for example see Babole (Leitch 2003), or in Eastern Bantu. Normally [ts] and [dz] can be written as <**ts**>and <**dz**> consistently within a language. In Francophone countries, with languages where [ts] and [tš] are contrastive, they are commonly written as <**ts**> and <**tch**>, as in French '*Tchad*' for English '*Chad*'. In addition, [ts] and [tš] can be dialectal variants within a language, and in this case the simpler [ts] <**ts**>can be used for both.

[ts]   Options: <**ts**> or <**tch**> (less desirable Francophone)

[dz]   Options: <**dz**>

Example : Dibole (C101) has forms like the following (Leitch 2000)

    <**dzálé**> *'open space in river'* /dzálé/ vs. <d**iàlè**> *'liver'* /di-àlè/

    <**edzóò**> *'evil tormenting spirit'* /e-dzóò/

    <**tsóó**> *'red, brown'* /tsóó/

    <**motsáàkò**> *'verbosity'* /mo-tsáàkò/

Voiced alveopalatal affricate [dʒ] and fricative [ʒ]

"Usually the affricate is more common than the fricative…so <**j**>is used for the affricate. However, symbolizing the fricative then becomes a problem (Gardner 2005:2)."

[dʒ]   Options: <**j**> (Swahili influence) or <**dj**> (French influence)

[ʒ]   Options: <**zh**> or <**jh**>

 <**zh**> makes a nice parallel with the voiceless fricative <**sh**>. "Other languages choose to reserve the <**j**> for the fricative and use <**dj**> for the affricate (e.g., Sena). However, this can end up introducing lots of extra **d**s into the written language (Gardner 2005:3)."

Implosives

Implosive bilabial and alveolar

Lubwisi, Bantu J, has an implosive b and d. The implosives are written by doubling the letters: <**bb**> and <**dd**>. Since Shona languages have both implosives and, rarely, explosives, they use <**b**> and <**d**> for marking the commoner implosives and <**bh**> and <**dh**> for the explosives, which are spoken with a bit of a murmur. This choice gives consistency of appearance with their murmured nasals, written <**mh**> and <**nh**>.

[ɓ]   Options: <**bb**> or <**bh**> or <**b**> (or <**'b**> or <**b'**>)

[ɗ]   Options: <**dd**> or <**dh**> or <**d**> (or <**'d**> or <**d'**>)

Using an apostrophe as part of a digraph is a bad idea. Don't do this unless you absolutely have no choice, i.e., you have a plosive voiced bilabial contrasting with both a fricative voiced bilabial *and* an implosive voiced bilabial. Even then, I recommend looking for a fricative symbol for /β/, such as <**v**>.

Double articulated stops

Labialized velars

[k͡p]   Options: <**kp**> or rarely <**pk**>

[g͡b]   Options: <**gb**> or rarely <**bg**>

Since the velar stop has usually been demonstrated to slightly precede the articulation of the bilabial, the digraph with the velar stop preceding the labial will usually be chosen over <**pk**> or <**bg**>, but there are exceptions such as Nyarwanda D61(J) and Shona (S10) (Maddieson 2003).

Geminate consonants

Ganda (E15) has a series of geminate consonants that are realized phonetically either by length or some sort of fortis articulation, according to "*A Primer on Speaking and Writing Luganda.*" The Baganda refer to this lengthening of consonants as "stress."

The Luganda consonants may be either stressed or hit softly. The letter <**b**> is a typical example of this. It is either strongly hit which requires the sound [**bb**] as in **mubbi** '*a thief*', or it may be hit softly as in <**mubi**> '*a bad person*'. The double consonant in a word is used to indicate a single strong emphasis on the consonant. This should not be confused with a repetitive consonant sound. Following are some more examples:

<**kutta**>   '*to kill*'   vs.   <**kuta**>   '*to release*'

<**mukka**>   '*smoke*'   vs.   <**muka**>   '*wife of*'

<**bidde**>   '*let them come back*'   vs.   <**bide**>   '*bells*'

The exception is the special consonant <**ny**>. When this is "stressed" in a word, only

the <n> part is written twice as in <nny> rather than <nyny>. One example is <nnyonnyola>.[22]

The formation of geminate consonants (a rare phenomenon) may indicate deletion of a high vowel from between them. Testing of an intelligent but naïve illiterate would be the best way to ascertain people's perception (see Gudschinsky 1973:132–133 for details).

Prenasalization

We could choose one nasal symbol, for instance <N>, to symbolize prenasalization everywhere. However, existing African orthographies write <m> before the prenasalized bilabials. Even in the case of palatal and velar nasals, a special letter for these consonants has never been used; rather, <nc> and <nj> are written for the palatals, <nk> and <ng> for the velars. Following are listed more problematic phones/phonemes.

Prenasalized labiodental fricatives

There is actually a distinction between the prenasalized <nf> and the heterorganic <mf> in some languages.

Examples from Rangi:

> <nfotera>   *'fold [it] for me'*
>
> <mfotera>   *'fold [it] for him'*

One may find prenasalized labiodental fricatives where a choice is possible between writing <nv> and <nf> versus <mv> and <mf>.

[^{m}f]   Options: <mf> or, less optimally, <nf>

[^{m}v]   Options: <mv> or, less optimally, <nv>

Most orthographies represent these as <mf> and <mv>, but some have opted for <nf> and <nv>. If the language also allows a *sequence of distinct* phonemes represented by <n> and <f>, for example, using this digraph will probably complicate recognition of phonemes in the language. That would be unusual, though.

Prenasalized palatals

Virtually all Bantu writing systems represent the voiced and voiceless prenasalized palatals as <nc> and <nj>, rather than something more graphemically complex, such as <nyc> or <nyj>.

[^{ɲ}t͡ʃ]   Options: <nc>

[^{ɲ}d͡ʒ]   Options: <nj>

Example from Tharaka:

> /^{ɲ}t͡ʃaana/   <ncaana> *'child'*

/$^{\widehat{n}}$dʒoka/ <**njoka**> '*snake*'

Prenasalized velars

According to Ron Moe (2002:8), "All existing Bantu orthographies write the prenasalized velars as <**nk**> and <**ng**>, rather than <**ng'k**> and <**ng'g**> or <**ñk**> and <**ñg**>. The 'correct' pronunciation is automatic on the part of the reader, of course, conditioned by the stop's point of articulation. As always, creating a trigraph with insertion of an apostrophe is not optimal."

[^{n}k]   Options: <**nk**> or <**ñk**>

[^{n}g]Options: <**ng**> or <**ñg**>

Prenasalized double articulated stops

Of course the only question is which nasal symbol to use, **m** or **n**, at the front of a trigraph. Usually **n** is chosen to represent the nasal.

[^{n} $^{\widehat{m}}$kp]   Options: <**nkp**>

[^{n} $^{\widehat{m}}$gb]   Options: <**ngb**>

Labialization and palatalization

Labialized consonants

[t^{w}]   Options: <**tw**> or <**tu**>

The orthographic challenge is choosing the spelling of a morpheme, [Cu] when another vowel follows the [u]. Syllable rules often determine the writing rule, converting the labial feature to <**w**>.

Example from Tharaka:

/mû + amba/ is written <**mwamba**> '*thief*'

/ba + kû + anda/ is written <**bakwanda**> '*planters*'

Palatalized consonants

[t^{j}]   Options: <**ti**> or <**ty**>

As with labialization, the choice of graphemes depends upon the structural (e.g., syllable structure) rules of the language and possible morpheme preservation issues. The same principle is in operation as the one for labialization, but the necessity of an underlying /u/ converting to a consonant /w/ when the underlying first vowel is high and rounded doesn't seem to be as strong for /i/ converting to a consonant when the first underlying vowel is high unrounded.

Example: The following chart of vowel sequences for Tharaka (Kithinji 1999:3-4), depicting the vowels in orthographic (not phonetic) form, illustrates this.

Figure 1 Tharaka Vowel Coalescence: when two vowels occur sequentially across morpheme boundaries

	i	î	e	a	o	û	u
i	ii	iî	ie	ia	io	iû	iu
î	ii	îî	ye	ya	yo	yû	yu
e	-	-	-	-	-	-	-
a	ai	ee	ee	aa	oo	aû	au
o	-	-	-	-	-	-	-
û	wi	wî	we	wa	wo	ûû	uu
u	ui	uî	ue	ua	uo	ûû	uu

The vowels listed down the left side are those which come first and those along the top are those which come second.

It is important to note that the vowel <**î**> only changes to <**y**> before <**e a o û u**> when <**î**> stands alone as a V syllable. Because of the language's phonological rules (syllable structure), when /I/ follows a consonant, as in <**kî-**> or <**rî-**>, it continues to be written as <**î**>and does not change to <**y**>.

/rû + a/ is commonly pronounced as *one syllable* (CSV), and is therefore spelled <**rwa**>.

/rî + a/ is commonly pronounced as *two syllables* (CV-V), and is therefore spelled <**rîa**>.

Palatal nasal [ɲ]

Most Bantu languages have a palatal nasal [ɲ] which is normally written with the digraph <**ny**>. The <**ny**> option corresponds well with Swahili and most, if not all, other Bantu orthographies. This does, however, conflict with the use of <**Cy**> to represent a consonant followed by a palatal. This is not a serious problem unless the palatal nasal [ɲ] is contrastive with palatalized alveolar nasal [nʲ](See examples below).

[ɲ]   Options: <**ny**> or <**ɲ**> or <**ñ**>

Many Bantu languages do make a distinction between /ɲ/ and /nʲ/, especially on the western side of Lake Victoria. Most eastern African languages write the palatal nasal <**ny**> as above, reserving <**ni**> for the palatalized nasal.[23]

Examples:

Zinza E23(J) in Tanzania with <**ny**> for palatal and <**ni**> for palatalized alveolar nasal.

Gungu E101(J) and Gwere E17(J) in Uganda use <**ny**> for palatal and <**ni**> for palatalized alveolar nasal (Oliver Stegen, pc).

---

[23]The palatal*ized* nasal is a phonetically complex phoneme. For more on this subject, see "Palatals" and "Palatalized nasals," The palatal nasal, on the other hand, is a single, simply articulated consonant.

Palatal nasal [ɲ] contrasting with palatalized alveolar nasal [nʲ] and palatalized palatal nasal [ɲʲ]

In some Bantu languages there is an emic contrast between the palatal nasal [ɲ] and the palatalized n [nʲ]. As mentioned above, Gungu has a three-way contrast between the palatal nasal [ɲ](<**kunyaala**> *'urinate'*), the palatalized nasal [nʲ] (<**kuhonia**> *'heal'*), and the palatalized *palatal* nasal [ɲʲ] (<**kukanyia**> *'cause to increase'* [Moe 2003:5]).

[ɲ]   Options: <**ny**> or <**ɲ**> or <**ñ**>

[nʲ]   Options: <**ni**> or <**ny**>

[ɲʲ]   Options: <**nyi**> or <**nyy**>[24]

There are problems with all of these options. Most orthographies use <**ny**> for the palatal nasal [ɲ], but this choice conflicts with representation of the palatalized alveolar nasal [nʲ]. If palatalization is normally transcribed as <**Cy**>, then <**ny**> will provide a consistent representation of that phoneme. But if both the palatal nasal and the palatalized alveolar nasal are contrastive emically and written identically, orthographic ambiguity will be the result for users of the writing system. The palatal nasal [ɲ] is by far the most common of the two. The palatalized alveolar nasal [nʲ] is most often written <**ni**>.

Sociolinguistically, mother-tongue speakers of Bantu languages have proven unwilling to write the distinction. The need for distinguishing the two may be compared to the need for distinguishing [ð] and [θ] in English—minimal pairs for them are rare indeed, with the voiced interdental fricative occurring in functors and the voiceless in content words, so they are both spelled <**th**>.

Test with connected text for comprehensibility and fluency of reading, as well as people's perception of the options.

Labialized and palatalized consonants contrasting with CSV sequences

Some languages permit a contrast between <**Cw**> (i.e., a labialized consonant such as /kw/) and <**Cuw**>. The Gungu words <**kwera**> *'yield'* and <**kuweera**> *'tell'* are illustrative. The same contrast can occur between <**Cy**> and <**Ciy**> (i.e., a palatalized consonant such as /py/ and a sequence such as /piy/, and between <**Cj**> and <**Cij**>. The Gungu words <**pyoko**> *'rhinoceros'* and <**rupiya**>, a loanword for *'money'*, demonstrate the same contrast. Other similar sequences are possible, of course.

[kʷe]   Options: <**kwe**> or <**kue**>

[kuwe]   Options: <**kuwe**> or <**kue**> or <**kuue**>

[pʲa]   Options: <**pya**> or <**pia**>

[pija]   Options: <**pya**> or <**piya**>, <**pja**> or <**pija**>

Regarding the last option listed, it may be inappropriate to use <**j**> to represent the palatalized voiceless bilabial stop, since it may already be used to represent a palatal stop, *not* palatalization. Choice of symbols depends upon whether the <**Cy**> symbol is already

---

[24] "The solution of writing a vowel here, at least for Lugungu, is a good one because the palatalization of the nasal is probably the result of a high [+ATR] vowel following the nasal" (Rod Casali, pc). If a high vowel is the cause of such a phenemonon, the vowel should be *written*.

in use, or whether, alternatively, <**j**> can have a dual role, representing both a voiced palatal stop/j] and palatalization of a voiceless stop <**pija**>.

Velarization

In addition to murmured consonants, all Shona (S10) languages have a set of 'labiovelarized' consonants (Gowlett 2003:614). According to Gowlett, "[the] *w* element carries both labialization in the form of lip-rounding, and velarization in the raising of the velum." With such an extensive and complex consonant inventory, digraphs and trigraphs are unavoidable.

Example from Shona:

"In addition to the basic phonemes, there are three major modifications: Prenasalization, Velarization, and Palatalization. Voiced oral stops and fricatives can be prenasalized (^{m}b, ^{n}d, ^{N}g, ^{M}v, ^{n}z, ^{n}z̧, ^{ɲ}j), represented by a preceding <**m**> or <**n**> in the orthography. Velarization, rather than labialization, is represented by a following <**w**> (e.g., <**bw**>, <**kw**>), and has several interesting phonetic realizations, as does Palatalization, represented by a following <**y**> in <**ty**>, <**dy**> and <**ndy**>" (Gardner 2001:59).

Shona, like many Mozambican languages, has an orthography that avoids the use of special characters. Besides the modifications mentioned above, the consonantal symbols in Table 1 are used (Gardner 2001:59). The large consonantal inventory requires many digraphs. Use of IPA symbols might be a better option for future orthographies, but LWC transfer issues are the challenge.

Table 1

Shona:	b	bh	d	dh	v	vh	sv	zv	sh	zh	ch	ny	n'	h
IPA:	ɓ	b	ɗ	d	ʋ	v	s̩	z̧	ʃ	ʒ	tʃ	ɲ	ŋ	ɦ

Morphophonological Processes

It's hard to keep this section distinct from the segmental items above it since most Bantu orthography issues have a morphophonological aspect! Donald Burquest (1993:172) says that there may be no differences between the representations of a structural phonologist, a generative phonologist, and a lexical phonologist.

This will *not* be the case for the Bantu languages. Morphology and phonology will impact one another repeatedly.[25] In the words of Keith Snider (2005:1),

> The 'trick' in developing a good orthography is to write words the way the mother-tongue speaker perceives a word to sound, not necessarily the way the mother-tongue speaker actually pronounces it. This is great for the immature reader. There is an added bonus, however, and that is that when we write this way, we are also able to maintain a constant word-image. Writing the way mother-tongue speakers perceive the language

---

[25]This is a very concise description of the issues. For more information, see Edmondson, et al. 1998 or Snider 2005. Snider's entire article can be found in Appendix D (Part 1).

> to sound meets the needs of both mature and immature readers with the same orthography.

At other times, however, the mother-tongue speaker is well aware of a difference in pronunciation of a morpheme. For examples of this in English, see *You can have your cake and eat it, too: An orthography that meets the needs of both mature and immature readers*, by Keith Snider (2005:3). in Appendix D (Part 1). The following are Bantu examples of morphophonological challenges and possible solutions.

Phonological conditioning at morpheme boundaries

Ron Moe, referring to hardening of voiced fricatives or liquids after a nasal, writes, "It would be possible to keep the underlying form of the root even when a nasal prefix is attached. For instance the /**r**/ in <**kuruga**> (root:-rug-) '*to come from*' could be maintained by writing <**nruga**>, which phonemically is [nduga] '*I come from*'. This would maintain a consistent morpheme image. However, most orthographies do not attempt to maintain a consistent morpheme image (certainly not in East Africa!). Some people have confused the principle of maintaining a consistent word image with maintaining a consistent morpheme image. Lexical phonology helps us to understand why the two are different… The results of lexical rules (such as the Gungu rule that changes the phoneme /r/ into the phoneme /d/ when a morpheme ending in **n** is prefixed) should be reflected in their spelling (Moe 2002:5). The result of such a lexical rule for Lugungu, E101(J), is <**nduga**>.

If allomorphs are morphologically conditioned they probably should be written reflecting their pronunciation, as above; if the variation is simply phonologically conditioned within the shape of the word, then the morpheme should be preserved in the spelling. In other words, show a change when morphemes come together. Don't show it when it's just the result of a phonological process such as compensatory lengthening.

Function words, especially short ones, are best written consistently, despite variation in their pronunciation. There is great advantage to having their form constant. Though they are often monosyllabic and therefore phonologically attached to a noun or verb, they should generally be written *disjunctively*[26] (see section on word breaks below).

The following example from the Lugwere writers' guide (Moe 2003:4)[27] illustrates phonological conditioning of a (word-internal) morpheme:

Option 1:

The prefix <**ki-**> 'Class 7' becomes <**ky-**> when it is followed by a vowel. However, the <**ky**> is pronounced the same as <**c**>. Both are pronounced [tʃ]. If a [tʃ] sound is from an underlying [ki-], it shall be spelled <**ky**>. Otherwise it shall be spelled <**c**>.

This option benefits fluent readers and fluency in general. The morphophonological alternation has effects beyond the word level, as can be readily seen by the matching prefixes in Option 1. The redundancy built into the spoken language is reflected in its written form, making words recognizable more quickly (with the letter <**k**> in the prefixes of both words, as in <**kintu kyange**>).[28]

---

[26]Literacy Note: they will need to be taught as sight words (memorized and recognized as whole words), but with a larger context whenever possible.

[27]The formatting of the original has been altered so that the symbols for phonemic, phonetic, and orthographic representations (< and >) are consistent in this manual.

[28]The <k> in the prefix of the first word tells the reader that the possessive pronoun following the noun

<**kintu kyange**>	*'my thing'*
<**kyama kyabwe**>	*'their secret'*
<**kinacegereekyo kikye**>	*'that trap of his'*
<**kintu kyaiswe**>	*'our thing'*
<**omu kyalo**>	*'at the village'*
<**kyama ekyo**>	*'that secret'*
<**kucicina**>	*'to giggle'*
<**ncaama**>	*'I stray'*
<**acuncumala**> (Lusoga)	*'he squats'*

Option 2:

The prefix <**ki**-> 'Class 7' becomes <**ky**-> when it is followed by a vowel. However the <**ky**> is pronounced the same as <**c**>. Both are pronounced [tʃ]. Whenever a word has the [tʃ] sound, it shall be spelled<**c**>, no matter whether it is from an underlying <**ki**->, or from an underlying <**c**>.

<**kintu cange**>	*'my thing'*
<**cama cabwe**>	*'their secret*
<**kinacegere eco kice**>	*'that trap is his'*
<**kintu caiswe**>	*'our thing'*

Option 2 has the advantage of one symbol, one sound, if after testing with careful, slow speech you find that people perceive the sounds to be identical ([tʃ] in both contexts).

Option 3:

<**ch**>, as in Swahili <**chumba**> *'room'*. Swahili may long ago have lost the underlying morpheme (noun prefix **ki-**) with the substitution of [tʃ] before some vowels in class 7, but is true to its perception by the mother-tongue speaker in its orthography, with a phonemic distinction between the variants.[29] The prefix is affricated at times, but not consistently: <**kiatu**> *'shoe' ,* <**kiazi**> *'potato',* <**kioo**> *'mirror'* **vs.** <**chakula**> *'food',* <**cheti**> *'paper',* <**choo**> *'toilet'.*

---

will of course begin with the same letter.

[29] This is described as a velar palatalization process in Nurse 2003:55.

Follow these steps to choose spellings for morphemes when phonological processes alter their pronunciation:

1. If the morphophonological process in question takes place across word boundaries, retain a constant visual image of the word. Associative markers, for example, are usually subject to phrase-level phonological changes rather than word-internal change, on the underlying vowel /a/, preceding vowel-initial nouns. Such grammatical words should have their integrity maintained: write morphophonemically across word boundaries (remember the "slow speech" principle).

2. Determine whether morphophonological processes take place between a root morpheme and its affixes.

3. If so, determine the following:

4. Do the changes affect the root morpheme?

5. Do they affect the affix?

6. Do they affect both root and affix?

7. If they do, determine whether the sounds which result from morphophonological processes are phonemes in their own right (phonemes which already exist in the language, and for which there is a symbol available). Try writing content words (nouns, verb roots, words with clear meanings) phonemically and grammatical particles morphophonemically.

Vowel coalescence/assimilation

Some vowels, originally distinct, change when they adjoin one another. Since these modifications are word-internal and involve changes into sounds which do exist elsewhere as phonemes in the language, reflecting the phonological changes in writing is usually best. Some morphemic contrasts are lost in this process, but beginning readers really benefit, especially for languages which have many vowels in the verbs. The vowel which is assimilated doesn't completely lose its influence; its influence upon syllable length often gives evidence of its presence in speech.

Figure 1 (see the chart of Tharaka vowel coalescence, or coalescent assimilation) illustrates the phonological processes which take place when different vowels adjoin one another and the orthography choices reflect them quite closely.

1. In general, spell words the way they sound *after* word rules have applied, but *before* phrase rules have applied (Snider 2001).

2. Determine whether the resulting sounds of the morphophonological process are sounds which may occur elsewhere in the language. If the sounds occur nowhere else, then they are conditioned allophones which should *not* have independent representation in the alphabet (Van Dyken, et al. 1993).[30]

---

[30]For more on this subject, see "How to make spelling choices" LinguaLinks version 5. [CD-ROM]. Dallas: SIL International, 2002. n. pag. Available: LinguaLinks Library bookshelf (see www.ethnologue.com/ LL_docs/lit_bkshlf.asp).

Principles of Tone Marking in an Orthography

Tone will be usually be underspecified in the orthographies of Bantu languages. You want to represent tone as simply as possible, while striving to maintain a constant word image to aid readers in morpheme recognition. Your main task here is to determine what must be disambiguated, i.e., to discover any contrasts which are conveyed *only* by tone. Here are some generalizations.

1. Expect not to mark surface tone. Expect to mark something more abstract, on the word level, if you mark tone. If you mark tone on a surface level, the reader will be forced to sound out every melody (i.e., sequence of tones on each syllable) every time he encounters a new word, in order to gain meaning.

2. It is the functional load (relative importance of a linguistic feature in conveying grammatical or lexical information) of lexical tone which determines whether or not it must be marked.[31] Is tone the only indicator of a grammatical or lexical distinction, or is there another indicator as well, within a word or phrase? If you consistently find different tone melodies across words of the same syllable pattern and same part of speech, it's a strong indication that lexical tone should be written. However, even in such a case, as few as one or two tone melodies may actually require marking (See the following Rangi example in "Nouns,").

3. "If there are more than 100 minimal pairs for tone in a given language, then writing of lexical tone should be considered. This is very unlikely for Narrow Bantu Languages. If one decides to write tone in a language, it is absolutely necessary to come up with a good way of teaching it", according to Kutsch-Lojenga (pc). If lexical tone must be marked[32], the most frequently occurring lexical tone is considered the default; hence it is not marked. For example, the root of a verb in a two-tone system could be marked by *nothing* if it is high and by a symbol over the appropriate stem syllable/s if it is low.

4. If both grammatical and lexical tone are found to carry a significant functional load, consider a double system of orthographic representation, representing both grammatical and lexical tone. In this way, the needs of both beginning and mature readers can be met. That is, you ensure consistency of writing, plus more immediate recognition of *meaning*.

   Steven Bird (1998:23) provides the following example from Kako (A93), a Bantu language from Cameroon. The injunctive form of the verb is indicated by a single diacritic at the end of the verb, while the tonal pattern of the word (conveying a lexical distinction) is indicated by a single diacritic above the first vowel of the verb stem.

---

[31]See "Functional Load" in the glossary, for a Lendu (Central Sudanic, nonBantu) example, in which tone has a very HIGH functional load. See also Appendix A, Determining Functional Load for Tone, for one procedure for calculating functional load.

[32]"In languages that have a tendency to be monosyllabic, there are numerous instances of tone distinguishing between lexical items or grammatical constructions. Bantu languages are [usually] on the other end of the scale. They tend to have long words, and tone is less frequently the only distinguishing feature between nouns especially. Tone plays a role which is much greater in the verbal system…there are not many instances where tone is the only feature distinguishing between two otherwise identical words" (Kutsch-Lojenga. 1986).

Figure 2

Orthography	Tone	Gloss	Orthographic Rule
a kêl!	à kêl	*he does-INJ*	circumflex=falling tone
a wôo!	à wóò	*he listens-INJ*	circumflex=high-low
a sânaŋgwɛ!	à sánàŋwɛ	*he works-INJ*	circumflex=high-low-low
a jêmbinaŋgwɛ!	àjémbìnàŋgwɛ	*he sings-INJ*	circumflex=high-low-low-low

Such a meaning-based representation of tone seems quite appropriate in a language where an entire melody can be compactly represented by one symbol per tone melody rather than one symbol per tone, or where tonally distinct words appear the same in isolation, or where there are morphemes manifested solely by perturbation of the tones of nearby words. In both situations, we may want to adopt diacritic symbols which are iconic representations of an underlying tone pattern (Bird 1998:23).

5. Along the same lines, for languages in which lexical tone need *not* be marked, but grammatical tone *does,* there are two possibilities. The first marks the tense at the beginning or end of a word. The second marks it directly over the contrastive morpheme.

    Mark a verbal morphemic distinction with a symbol at the front of the verb, such as this. The value of this marking system is that the reader/writer simply knows that this tense, i.e., distant past, is indicated by that symbol. No awareness of the locus of an affix or a syllable to which tone is attached is necessary.

    In this example, Kwaya speakers chose to distinguish between present perfect completive tense and the distant past tense (2 days ago and before). Therefore, the distant past tense will be marked with a word-initial colon.

    <naateekere>   'I have already cooked'

    <:naateekere>   'I cooked (two days ago)'

Snider suggests positioning the diacritic in some consistent position directly over the morpheme which would otherwise be ambiguous.

Diacritic marking of grammatical tone can be successfully employed when tone does not play a great role in distinguishing lexical items, but there are nevertheless a significant number of grammatical distinctions indicated by tonal differences. In this case, one does not mark tone per se. Instead, one uses diacritics in a consistent manner to indicate the different grammatical distinctions. An acute accent on the verb, for instance, might be used to signal perfective aspect, regardless of how tone is realized in this construction. Similarly, a circumflex might indicate imperfective aspect (Snider 1992:7).

6. The best way to discover what additional grammatical tonal phenomena must be marked is for the local speakers to start using the fledgling orthography and keep track of any cases of sentence-level ambiguity (i.e., a sentence may be written according to orthographic rules but still have two different interpretations because of unspecified grammatical tone phenomena).

7. Floating tones: If floating tones are contrastive, represent words that have floating tones differently from words that do not have them. In other words, mark the difference on the word that *causes* the difference, as opposed to the adjacent word that undergoes the change.[33] If we need to represent each tone melody in a unique fashion anyway, and since the floating tone of a particular word is part of its inherent melody, it only stands to reason that the word that initiates the floating tone should be the one that gets the special treatment. (For more examples and explanation, see Snider 1998. SIL Africa Area has a copy, on CD.)

Verbs

8. Determine the functional load of the verb tone contrasts (this is after contrastive melodies have been identified during phonological analysis). Many Bantu languages will have two or occasionally three tone classes (underlying pitch patterns) for verb roots.[34] In others, there is no longer a lexical tonal contrast. If there is more than one tonal melody for verb roots, test readers for the need to mark tone on the verb root. Minimal pairs are just a *clue* to possible heavy functional load, not the only indicator. Give the verbs a context when testing.

9. Verbal extensions: Verbal extensions not only have a reduced vowel system—they usually also lack tonal distinctiveness.[35] Don't expect to write any lexical tone on verbal extensions because they don't normally have underlying tone. A grammatical tonal pattern, however, will be realized across the whole verb, including any extensions.

10. There will probably be some Bantu languages, particularly in the west, which require marking of verbal tone both lexically and grammatically. See point 4 above for details on using a dual marking system which indicates both lexical and grammatical tone. Determination of the necessity for such dual markings may require testing.

    I recommend using the two distinct kinds of marking whenever both lexical and grammatical tone must be written: marking lexical classes of verb or noun roots tonally by the use of a diacritic over a vowel (if tone class marking carries a high functional load for that language) and marking the grammatical tonal melody of a specific grammatical verb form with an abstract symbol, such as an exclamation

---

[33] A floating tone (generally a L tone) causes a following H tone to be realized at a lower pitch than it would be if the floating L tone were not there.

[34] Proto-Bantu had two lexical tone classes for verbs, and they still exist in the great majority of Bantu languages (Maddieson 2003:40). Narrow Bantu languages often have a third exceptional class, referred to as toneless verbs. This class of toneless verbs adopts its surface tones from its surroundings or from adjacent tonal morphemes. On the surface a toneless verb may appear to function at times like a high tone verb and at other times like a low tone verb. This is most often the case with very common, and hence often irregular, verbs.

[35] Schadeberg (2003:148) says, "In the verbal word, they generally adopt the (first) tone of the final suffix ('tonal extension assimilation' by anticipation)."

mark, colon, hash, or other mark. The use of such a meaning-based symbol at the front of the verb alerts the reader to the correct tonal melody of the word. Such a symbol will not be taught as a mark of pitch/tone, but as a grammatical marker (i.e., # equals recent past tense), and should be recognized like a sight word. This amounts to morphemic writing of grammatical tone (Bird 1998:16).

In languages with shorter words, distinctive markings directly over the vowel may still be the easiest. Many Grassfields Bantu languages use diacritics for grammatical tone. According to Anderson and Alomotor (2005), Awing requires the marking of habitual aspect. Diacritics are used for this purpose, with two dots over he vowel <ä>. The rule is that this diacritic be placed on the first vowel preceding the verb in a sentence. Note the contrast in the following example sentences.

Nouns

11. Try to find a system to distinguish tone for nouns if lexical tone proves to make a distinction for several minimal tone pairs (e.g., 5 percent or more of words in the lexicon). Use the pitch pattern of the noun in isolation, unless for some reason underlying pitch patterns are neutralized in isolation. Expect to find at least two main pitch patterns for nouns; many languages will have more, such as Bila in Eastern DRC and Tharaka in Kenya.[36]

Functors

12. Marking tone in functors (small grammatical words such as *to* and *too*) can be very important, if sentence order does not already tell the reader which grammatical word they are reading. Most important is to distinguish functor minimal tone pairs that can be confused in relatively short sentences.

Word Boundary Principles

Word breaks have always been a challenge in writing Bantu languages, partly because of the extensive affixation of monosyllabic morphemes.

We can assist fluent word recognition by matching the intuitive knowledge of mother-tongue speakers. The following guidelines should help you do this (Van Dyken, et al. 1993:7). Examples added are mine.

1. The principle of **referential independence**: A morpheme qualifies as a word if it clearly communicates meaning, even when heard or seen in isolation. Conversely, each language has certain morphemes which cannot communicate meaning in isolation, such as Bantu noun class markers.

2. The principle of **conceptual unity**: This requires that each written word convey only one concept. When compound words are formed, it is because the two words join to form a new and unique concept.

---

[36] Proto-Bantu had four distinctive tone patterns for disyllabic nouns: LL, LH, HL, and HH, which might be expected in most Bantu languages; even reduced tone systems with only two or three lexical tone distinctions on nouns might still require tone-marking. Tharaka has an extra-high tone.

3. The principle of **minimal ambiguity**: Word space can show that two words are pronounced with different stress patterns (e.g., *blackbird* and *black bird* are distinguished orally by different stress patterns). Ignoring these differences can lead to ambiguities in the written text which are absent in speech. If both *blackbird* and *black bird* were written with a space between the two component words, the difference in the two meanings would be lost.

    Example: In Tharaka, the proclitic /wa/ '*each*' has an entirely different function (and stress pattern) from the /wa/ written separately, meaning possession, so the writing system reflects the distinction.

4. When a morpheme demonstrates **mobility**, it is usually written as a separate word, even if it doesn't make sense in isolation. In Swahili, **Wageni wenu watafika lini?** Or **Lini wageni wenu watafika?** are both correct (Translation: *When will your guests arrive?*). Since the adverb **lini** clearly is not obliged to follow the verb and its affixes in this example (**watafika**), but is movable (it is isolatable as well), it should be written as a separate word from the verb, for readability considerations.[37]

5. The principle of **separability:** A functor is written separately from the word with which it usually occurs when other words can intervene *between* them. In Bantu languages the subject prefix, negative marker, and tense and aspect markers tend to be bound affixes; because these markers are so closely linked to the verb stem there is no possibility of inserting any other words between them and the verb stem.

    Example: The associative marker, usually consisting of a prefix plus –a, e.g., the /ya/ in Swahili **nyumba ya mdogo wangu**, '*the house of my younger relative*', should be written either as a clitic or as a completely separate word from the following noun. Other words may intervene between the associative and the noun, as in this example: **nyumba ya huyu mdogo wangu** '*the house of this younger relative of mine*'. The associative should not be attached to a noun as a prefix, because it is a separate word. Since it is monosyllabic, however, people may prefer to treat it as a clitic. (See following section on clitics.)

    However, in some verb forms, the prefixes may be bound to an auxiliary, while the main verb stem constitutes a separate unit. If it is possible to insert other words between the associative marker and the verb stem, it is quite possible that there is a verbal auxiliary, and thus there is a word break in the verb. In this case, people will respond favorably to writing the verb stem as a separate unit. The subject marker will probably be perceived as a prefix on the word which precedes the main verb stem within the verb phrase. Phonological clues such as vowel harmony, vowel length, stress, and tone may help guide the word break decision. If not, test speakers with a variety of treatments of word breaks in such verbs.

---

[37] Based upon its mobility, a poor word break choice for Swahili's interrogative particle /je/ is shown here. It can occur in two locations in a sentence: <Je, wamefanya nini?> 'What? What have you been doing?' and <wamefanyaje?>'You have been doing what?' It should be written disjunctively in *both* instances, though the stress patterns of speech have supported its conjunctive spelling.

Tone can also make a difference in grammatical meaning, and of course that written distinction must be attached to the part of speech it modifies.

6. The principle of **substitutability** is often a criterion for writing certain grammatical elements as separate words. In Bantu languages, one can substitute 'the boys' for the independent pronoun ' they' in < **wao walicheza** > *'they played'* or < **wavulana walicheza** > *'the boys played'* but not for the same prefix in the verb. One cannot say, for example, */wavulanalicheza/. One must say /wavulana walicheza/. Hence the /wa-/ morpheme is written as part of the word.

As stated earlier, there may be much elision in rapid speech. When at all possible, choose the full form with the different part of speech as a *separate word,* striving for the constant word image. Good readers always recognize units which are larger than the individual grapheme; they quickly recognize and retain syllables and morphemic units. Remember (at the risk of overstating it), spell words the way they sound after word rules have applied, but before phrase rules have applied (Snider 2001:324).

A linguistic consultant will take advantage of phonological indicators such as tone patterns which can only occur at certain places within a word, vowel length indicators, and vowel coalescence indicators, in finding word breaks. Coalescence of vowels within a word results in a different vowel quality than coalescence of those same vowels across a word boundary, following phrase-level rules. These patterns, phrase-level and word level rules, can really help one distinguish word breaks, and mother-tongue speakers will probably quickly identify them. All of these indicia should be used to inform orthography development. The following issues are thornier and very common.

Clitics vs. affixes: two different animals

One major cause of confusion plaguing the issue of word division in Bantu stems from the failure to distinguish between affixation and cliticization. The failure to distinguish between monosyllabic affixes, which are an integral part of a word, and monosyllabic words which are separate words, phonologically cliticized to another whole word, is a major cause of the confusion regarding word boundaries in Bantu.

Clitics are phonologically joined to a host word and yet are not affixes. Whether they need to be represented as separate words or as words attached, such as by hyphen or apostrophe, to another word depends on whether or not this joining of the clitic to the host word causes phonological changes such as tone or stress changes, or changes in vowel length or quality. In order to preserve the word shape both of the clitic and the host word, it is not recommended to represent a clitic in the same way that one represents an affix.

Clitics are a common occurrence in Bantu nominal and verbal phrases. Mother-tongue speakers, especially those with less exposure to print, will often want to see them written conjunctively with the word on which they depend because they are phonologically connected. A writing system must do more than represent *sounds*. It represents the function and meaning of sounds and morphemes to the reader. The big challenge with clitics is representing their separateness, yet still acknowledging their phonological attachment to the word. The big question to ask is: "Is it an affix, or is it a clitic?"

This appears to be a simple question, yet depending on the morpheme involved, it may or may *not* be that simple! The six principles found above under "Word Boundary Principles"

can help you determine if a morpheme is a clitic or an affix. Once the question is answered, these steps are suggested by Karen Van Otterloo (pc):

> If it is an affix, it should be written as part of the word.

> If it is a clitic, the OWL should decide whether there are any phonological consequences for the host word. Is there a change in stress? Does vowel lengthening result? Are there tonal changes in the host word? If the answer to any of these three questions is "yes," it will probably be best to indicate orthographically the clitic's dependency on the host word (e.g., with hyphenation).

> If the answer is "no"—there are no such consequences of cliticization upon host words in the language—then a literacy consultant should determine the mother-tongue speaker reaction to writing the clitic separately or together. If mother-tongue speakers choose writing the clitic conjunctively, the next decision whether to use a hyphen to mark the joining.

A useful generalization is that locatives attached to nouns will usually be clitics. Corresponding locative concords, used with verbs, pronouns, adjectives, and demonstratives, will be affixes.

As long as this difference between the nominal and nonnominal elements matches the perception of mother-tongue speakers, there will be no problem in getting them to write the locative plus noun as a clitic (or separate word) plus a noun, while writing the concord prefix plus verb as a single word.

As we've said, Bantu languages have strong prohibitions against monosyllabic words, thus pressuring such clitics as the Tharaka /-ni/ '*in, on*' to be written as part of the nouns they follow. Try a hyphen to show the separateness, yet phonological and semantic dependency of such a locative.[38] Tharaka /tûragita≠ni/ '*on the tractor*' would be written <**tûragita-ni**> rather than <**tûragitaani**>. The relative separateness of the locative even helps to visually preserve the appearance of the word <**tûragita**> since the slight lengthening of /a/ is probably due to the fact that it is now penultimate.[39] If there are other enclitics in the language, they should probably be written in the same way.

Options: [-] or [']

Hyphenation is one possible indicator of that close relationship. It shows dependence yet distinction. Another is use of the apostrophe.

The rationale for using an apostrophe is that with cliticization, there is almost always vowel elision or coalescence taking place. The apostrophe does double duty in such cases, signaling to readers that it is tied in with the following word and automatically allowing for any elided vowels (Karen Van Otterloo, pc). In Francophone countries, the symbol is usually well-received.

---

[38]As further support for writing locatives as clitics, Bresnan and Mchombo (1995) have shown that in the various Bantu languages they have studied, locative markers used with nouns are not prefixes, while the nonlocative noun class markers are.

[39]Another indication of the distinctiveness of the Tharaka locative is the fact that the tone of the word-final vowel on the noun always changes preceding the locative. As Karen Van Otterloo (pc) says, "If it adds a new element to the melody, it is probably a clitic."

A *minus* to use of the apostrophe is that in some orthographies the apostrophe has other functions, as in Swahili, where it is actually part of a trigraph <**ng'**>. It's best not to use the symbol to represent such a diversity of linguistic phenomena. A typical mother-tongue English speaker, highly literate, struggles with confusion as to whether to spell the possessive *its* with an apostrophe as *it's* (incorrect), as is done with the possessive, e.g., <**dog's**>or the contraction for *it is*, <**it's**>. This recurring problem is due to the multiple functions of the apostrophe and the multiple functions of /-s/ both in the spoken language and in the English orthography.

If, however, mother-tongue Bantu speakers choose to use <**ŋ**> to represent the velar nasal in their orthography, overuse of the apostrophe is not a problem!

Augment vowels

Augment vowels (e.g., a, e, o) sometimes come before the noun class markers, e.g., /ama-/, /omu-/, and /eki-/.

Many Bantu languages have, in addition to their noun class prefixes, an augment or pre-prefix, which in some languages functions as a definite marker. Swahili does not use the augment, so we will exemplify it by using Fuliiru D63(J), a language of eastern Congo. Fuliiru's augment[40] appears as a single vowel preceding the noun class prefix, e.g., u-mu-ndu '*person*' a-ba-ndu '*people*' and i-ki-ndu '*thing*'. This pre-prefix is generally not separated from the noun (Van Dyken, et al. 1993:10).

There are some languages in which the pre-prefix functions as a clitic rather than a prefix, however. In such cases, it may best be written as such. To write it as a completely separate word is not a good option because the form of the vowel is dependent on the prefix that follows, and in some cases the length of the vowel is also dependent on what follows it.

The presence of augments is determined by different factors in different languages, but they typically carry minimal semantic information. Since they always occur word initially

---

[40]According to Van Otterloo (To appear), in Fuliiru a noun class prefix is always an affix. Locative markers, on the other hand, while often presented in the literature as unquestionably "noun prefixes," function syntactically, at least in Fuliiru, as monosyllabic words rather than as prefixes. If we compare, for example, the class 19 diminutive nominal prefix, with any locative marker, we see that there is an essential difference between the two in terms of separability. The diminutive prefix usually replaces the normal noun class prefix of a noun which is being diminutivized. Especially in class 9/10 nouns, however, the diminutive is sometimes added to a complete noun, preceding its normal noun class prefix. Even when added to a complete word, however, it is still clearly an affix and not a clitic, because it is inseparable from the noun to which it is attached. Compare ny-úùmbà *9-house* and its diminutive, **hínyúùmbà** *19-9-small house*, with the phrase ìyó gîìndì nyúùmbà *that other house* and its diminutive counterpart yìhyó hîìndì hínyúùmbà that *other small house*. Note that the hí- prefix is never separated from the noun to which it is attached. Rather, it remains attached to the noun, and the modifying words within the phrase have matching concord prefixes.

A locative marker of any class, on the other hand, also precedes a whole noun, e.g., ny-úùmbà *9-house/* mú'ny-úùmbà *in the house* (18-9-house). However, the locative marker, unlike the diminutive prefix, *is* separable from the word to which it is phonologically attached. Observe how the locative marker appears only at the left of the phrase as modifying words are added: mú'gîìndì nyúúmbà *in another house*, mwìì'yó gîìndì nyúùmbà *in that other house*. Clearly, the mú is independent from any specific part of speech, and rather is joined phonologically to whatever word which happens to follow it within the phrase. Thus, the locative marker *can* be separated from the noun to which it seemed to be attached in the original phrase mú'nyúùmbà *in the house*. It simply remains the head of the phrase as other components are added, and there is no concord within the noun phrase that agrees with the class 18 locative marker.

if they occur at all, they invariably interact with the monosyllabic and therefore cliticized associative markers, e.g., **bya**, **za** *'of'*, coming before them in an associative phrase. For example, in Fuliiru /i-bi-ndu by-a u-mw-ami/ becomes [ibindu byomwami] *'the things of the king'*, according to Karen Van Otterloo (pc). The augment itself may be an affix or a clitic, depending on the particular language. This issue can probably be determined by following the word break criteria given by Kutsch-Lojenga (1993:7), applying the principles of mobility, movability, separability, conceptual unity, and referential independence. Consider, for example, whether it is possible to insert another word between the augment and the noun it precedes, or whether it is possible in any circumstances to pronounce the noun in a sentence without the augment.

The best option for longterm readability is preserving the underlying morpheme, <bya umwani> in this example, though speakers may want to write it as it sounds as part of the phrase.

Options: < #_> or <'> or <-> or <**word break**>

The first option is affixation. In the Malila and Shi examples which follow, the augment vowel should be written together with the rest of the word. Kutsch-Lojenga (pc) says, "the quality of the vowel *depends* on the vowel in the prefix which follows (in Malila it must be the same as the prefix vowel; in Mashi, it is e, a, o, according to the front, central, back vowel of the prefix); often (and certainly in Malila), the whole word, including the augment, forms a *tonal unit*." Tonally, augment + prefix + noun form a unit.

Malila (M24) examples with augments:

ú-mu-tiinho	*ladle*	**i-shi-péeni**	*knife*
ʉ-mʉ-líindʉ	*girl*	**i-shɨ-bhaanza**	*congregations*
a-bha- líindʉ	*girls*	**ú-lu-kusa**	*rope*
á-ma-fupa	*bones*	**ʉ-lʉ-khole**	*muscle*

For all of them, the following orthography was chosen: <**úmutiinho**>, and not <**úa mutiinho**>; <**ʉmʉlíindʉ**>, and not <**ʉ mʉlíindʉ**>.

In the following examples from Shi (D50), the vowel of the augment agrees in front/central/ or backness with the prefix vowel. For front and back, it is one degree lower than the prefix vowel (Kutsch-Lojenga pc). Tone is not marked in these examples.

bantu	*people (indefinite)*
abantu	*people (definite)* not: a bantu
rhubwa	*dogs (indefinite)*
orhubwa	*dogs (definite)*
mirhi	*trees (indefinite)*
emirhi	*trees (definite)*

The second possibility is that the augment is a clitic. An indication of semantic separate-ness (though phonological attachment) of a clitic is vowel lengthening. According to Karen Van Otterloo (pc), if the vowel of a noun prefix is not lengthened when the noun stem is monosyllabic, but the vowel of the augment is lengthened when the noun is short, this can be one indication that the augment is a clitic rather than a prefix. In Fuliiru an augment undergoes lengthening in such a situation, and because of its mobility and separability it is clearly identifiable as a phrase-level clitic. The augment, in such a case, can be considered a monosyllabic word which attaches phonologically to the following noun or adjective. The following examples are from Fuliiru (Van Otterloo, to appear).

	UF	**Phonetic**	**Gloss**	**Orthographic form**
a)	í kí-he	[ííkíhe]	*AU'7-time*	< ikihe >
b)	í kí-nògòsho	[íkínògòʃò]	*AU'7-hoof*	<ikinogosho>
c)	á bá-nà	[ááβánà]	*AU'2-four*	<abana>
d)	á bá-àná	[áβāānà]	*AU'2-child*	<abaana>

An augment, then, may be a monosyllabic word, and as such *may* be written separately.[41]

Associative marker (wa, ba, gwa, gya, etc.)

Associative markers should usually be written separately from the nouns to which they refer. A word break challenge arises when the word following the associative begins with a vowel (often an augment). This vowel often alters the pronunciation of the associative marker when it is part of a noun phrase, because the vowel of the augment typically harmonizes with the vowel in the class prefix (Gardner 2005:30) in the word which comes after it.

The following example is from Ikiizu (E402).

1/1a	umuryakari *bride*	wa *of*	baana *children*
2/2a	abaryakari *brides*	be *of*	kesebe *udder*

---

[41]Literacy specialists seem to disagree with linguists on this point. Literacy people hope to facilitate reading fluency and comprehension by indicating the semantic separateness of the distinct words. They hope readers will eventually be able to read *silently*, because comprehension is much better with silent reading; the reader is not forced to simultaneously decode print, vocalize, and listen with comprehension to his own speech.

3	umuri *root*	gwi *of*	nyumba *house*
4	imiri *roots*	jɨ *of*	ngiri-gwasi *warthog*

The associative has an underlying <**a**> as its vowel, but because it is pronounced together with the following word, the <**a**> coalesces with the augment and elides. For more detail, see section on affixes vs. clitics.

Since word level rules, not phrase-level rules, usually guide spelling, it is best both for writers and for readers if associatives are always written with <**-a**> word finally. When associatives precede vowel-initial nouns, mother-tongue speakers may want to write them reflecting the vowel coalescence processes of their language. This makes the associative marker difficult for the reader to recognize because its form will vary enormously, and because the one unvarying letter before all consonant initial words, <**-a**>, will be missing. Such surface-level writing, though it helps new readers, eventually slows spelling and writing rules. Acceptability to mother-tongue speakers must be considered, however, so present the options carefully (see Part 2, appendix C).

Options: <**-a#** or **–V#V**>

In some languages, e.g., Babole (Leitch 2003), there may be a difference between associatives used for possession and those used for all other purposes. For possession there is an "extra" initial vowel that is absent in the other associative forms. This is seen in the contrast between (1) and (2) below.

(1)   <**byèkà   byá   bààbútì**>

        8:food   8:ASS   2:visitor

        '*food (set apart) for the visitors*'

(2)   <**byèkà   íbyá   àmɛ́**>

        8:food   8:ASS   me

        '*my food*'

The additional morpheme -í- in (2) looks like a remnant of an 'augment' and only occurs in possessive associatives (as opposed to generic associatives). In rapid speech, this initial vowel commonly elides, and is easy to miss, but the distinction, if it exists, must be represented in the orthography. For additional details see Leitch 2003:419.

Options: <**word break before noun/pronoun with underlying -a preserved in writing**> or
             <**word break before noun/pronoun with changes in pronunciation written**>

Options for writing them do not include hyphenation. Associative markers should not look like locatives, stuck onto the beginning of the noun. Also, since in Swahili the associative is written disjunctively, people should not object to writing them separately from the nouns they precede. People do not usually question the suggestion. A problem arises, however, in that many people will want to follow phrase-level, rather than word level (lexical rules), as pronounced when the associative is part of a noun phrase in their speech. For example, an associative marker which precedes a vowel-initial noun will often have its pronunciation altered by the vowel of the noun it modifies, in rapid speech, or by vowel harmony rules.

In the following example from Ngoreme, note the changes in the vowel of the associative marker for some noun class 1/1a words:

| 1/1a | **omugabo** | **wo** | **moona** | **wa** | **baana** | **we** | **nyumba** |

The Ikiizu (E402), on the other hand, chose to just write the associatives and the locatives in their underlying forms. After all, they have seven vowels, and the variant speech forms are many indeed. They also are convinced that the associative is always "wa," despite its morphing in speech before every V-initial noun.

The conjunction na

This is a nonagreeing associative marker, according to Van Otterloo (pc). As such, it should be written separately, or as a clitic attached to the word it precedes with <'> or <->. A word break will facilitate recognition of the word which follows it and will also avoid merging different parts of speech into one written form. If a word break is not possible, try indicating its separateness with an apostrophe or a hyphen.

[na =]   Options: < **word break** > or <'> or <->

Vowel elision/coalescence in clitics

As noted earlier, much elision and coalescence of vowels is done at phrase-level in rapid or normal speed speech, and since these vowel changes result from phrase-level rules rather than word level rules, they should not be represented in the orthography. However, with clitics, there may be undoable elision, even with unnaturally slow speech. Especially in the case of augments which are clitics rather than prefixes, there may be word level coalescence of vowels between the augment and another preceding clitic such as the associative marker. Since such elision/coalescence is not truly a phrase-level rule, it may best be indicated orthographically.

In determining whether the augment must be written on the noun after an associative or conjunction, try to find out two things:

1.   if there are any instances when the augment cannot be pronounced at the beginning of the noun, because it changes the meaning when it is pronounced together with the noun.

2.   if there are two ways to pronounce the augment in the same phrase (one reflecting fast speech and one preserving the underlying forms).

If either of these conditions is true of the language, write the augment in its full form, and write the associative in its underlying form, usually <Ca>. No elision! Readers will

more quickly recognize the associative and the noun which follows it, and within a few weeks of practice they'll read aloud as they actually speak.

If possible, determine a base form for each clitic and maintain that shape (spelling). This will help develop fluent reading, but in the early stages of literacy, it will make the teaching of writing more difficult if it is conflicting with word level rules.

If the coalescence involves grammatical words which cannot be isolated from another word even at the slowest rates of speech, this indicates a word level rule; the spelling should indicate the functor as part of the word. The solution is to reflect linguistic reality. If the augment is actually a clitic, it should be written as such.

> The noun plus augment would appear orthographically as a clitic group: e.g., u'mundu. In such a case, the only option for the conjunction plus noun, phonetically [nomundu] would be to write: <**no'mundu**>. This would accurately reflect the linguistic facts that the augment is a clitic and that word-level coalescence takes place between the conjunction and the augment. This would maintain the word shape of the noun stem in both cases. And while such a solution would result in several shapes for the conjunction no', ne', na', etc., (e.g.,< **i'kindu**> *the thing,* <**ne'kindu**> *and the thing,* <**a'bandu**> *the people,* <**na'bandu**> *and the people*), this different shape would still accurately reflect the shape of the conjunction at word-level. The same sort of allomorphy occurs in noun prefixes at word level, and no one would even think of trying to undo it for the sake of maintaining word shape. For example, in Swahili, the class 2 prefix wa- may appear as <**wa-**> as in watu people, or as <**w**>- as in <**wote**> *all* or<**wema**> *good* and no one would think of writing <**waote**> or <**waema**> just to keep the morpheme shape the same (Karen Van Otterloo, pc).

In summary, aim to reflect the linguistic status of the lexical units (the "linguistic reality"). Also, try to reflect word level rules, not phrase-level rules. The best result of an orthography decision for a clitic is one which allows people's silent reading to be fluent (recognizing meaning quickly) and allows their oral reading to make adjustments, so they can pronounce a word as it would be spoken in rapid speech. People can usually be given some linguistically acceptable options, and then are able to choose between them, depending upon how they want to indicate the coalescence.

In verbs

Bantu verbs have so much affixation that people tend to think that everything preceding the verb stem must be a prefix. As Karen Van Otterloo (pc) points out, however, there are usually:

> …at least one or two verb forms in which there is an auxiliary (or even multiple auxiliaries in a single verb, cf. English *will have been gone,* and Kifuliiru <**bagweti bagashubi yimba**>'*they are intentionally singing again*'). In verbs with auxiliaries, each auxiliary is a separate word from the main verb. The main verb is *gone* in the English example, and **yimba** '*sing*' in the Kifuliiru example. The most common separate-word auxiliary will probably involve a form of the verb '*to be*', often **–li/-ri/-ni** or **–ba** , which will bear the subject prefix, perhaps along with other[s].

The main verb which follows may be infinitival, or it may also have a subject or object prefix, or even be a bare stem with no prefixes.

Generalizations

Expect a copula which is followed by a main verb to be written separately.

Expect to find two separate verbs whenever you encounter two subject prefixes in a stream of speech, as in the Fuliiru sentence /bagwetibagashubiyimba/ '*they are intentionally singing again*'; a natural word break will be signaled by a subject prefix, so <**bagweti bagashubiyimba**> will be the first attempt at word breaks. In Fuliiru there are other indicators which show that there is also a word break between /bagashubi/ and the main verb /yimba/, thus giving <**bagweti bagashubi yimba**>, a three-word verb form (Karen Van Otterloo, pc).

Expect the presence of an infinitive marker to also indicate the beginning of a separate verb. You may find either the normal infinitive marker **(ku-)** or a prefix of noun class 5, which is usually an /i-/ or an /e-/.

Expect to find a subjunctive final vowel (usually /-e/) on the main verb of a subjunctive verb form. If the subjunctive ending has instead been replaced by a default final vowel, this may indicate there is a word break in the verb. Often the subjunctive ending will not show if there is an auxiliary between the prefixes and the main verb.

Options: <**word break**> or <**-**> or <**'**>

As we said when we began, linguistic analysis will undergird any writing system which accurately represents the speech and perception of the mother-tongue speaker. Toward that same end, the principles and suggestions given in this chapter incorporated pedagogical, perceptual, and sociolinguistic considerations into the decision-making process for orthographic choices. Your team should now be prepared to make word break and grapheme choices, which will be recorded in the Workbook Form. Most tables and charts in the Workbook Form are set up to hyperlink the reader back to the linguistic features described above.

Chapter 3 - Orthography Testing

Troubleshooting Problem Areas

In the words of Keith Snider, "As with everything else in orthography design, try something and then test it thoroughly" (Snider 2001:6). If you know what to look for and how to identify it, you will be able to assess the **accuracy** of the linguistic analysis, the social **acceptability** of the writing system, its **readability,** and even **writeability**.

Usually, two aspects of orthography **readability** require testing: comprehension and fluency. Comprehension can be tested by having subjects read one or more selections and then asking content questions. Fluency can be tested by having subjects read one or more selections orally and then noting the places where they stumble. The people who conduct the test should record the different readings on audio cassette, and then later note the number and type of reading errors made.

A note on the value of tape-recorded tests: A literacy specialist who is not a mother-tongue speaker of the language is likely to miss people's mistakes with oral reading. He/she will be limited in distinguishing some minimal pairs, for example, especially in the flow of speech in a text. Since the passage can only be read *once* by each reader, there are no second chances for the *tester* to note everything! If the test is recorded, though, you can listen repeatedly for problems, along with a mother-tongue speaker or two.

You may also need to test people's ability to **write** the language, applying its spelling rules. There are also two aspects that should be tested: accuracy and speed. A dictation test usually suffices, with the tester noting the number and type of errors made. If people are very slow in writing the words dictated, the tester should note this as well.

Probable Test Needs

1. Speakers' perception of a feature of their language which usually has lower speaker awareness (i.e., breathiness, tone, length)

2. Aspects of the language which are left ambiguous for the reader—level of difficulty in guessing whatever is not specified or contrasted

3. Word break (and comprehensibility) questions with long words or clitics

4. Vowel deletion when it comes across word breaks (i.e., Is it possible to write out both words fully?)

5. Vowel coalescence and morpheme loss or other loss of contrast

6. Spelling rules—which, if any, confuse people?

7. The extent to which tonal contrasts need to be marked

8. Mother-tongue speakers' response to altermother-tongue representations of tone

Who Should Be Tested

This depends upon the purpose of the test. Usually, though, test:

1. Members of different dialect areas

2. People who are literate in a language of wider communication

Types of Tests

Informal vs. Formal

Informal

1. Just ask people if they have difficulty reading or writing the same feature in a variety of contexts.
2. Elicit mother-tongue speaker reaction to a problem area. Asking people "Which option do you prefer?" predisposes them to choose whatever they've been accustomed to seeing in another language. Instead, say, "Which option is true of your language?"
3. Observe people trying to read or having problems in literacy classes.

Are certain symbol sequences difficult to teach?

Are students having difficulty spelling certain symbol sequences?

Formal

1. Define the problem to be tested.
2. Design a test which will give you information on how to solve the problem.
3. Decide whether the test should be oral or written. You can ask people to read aloud, to write what is read to them, or to read silently and express their comprehension of what was read (i.e., a minimal pair).
4. Carry out the test.

Giving formal tests

Oral test with a text

1. Ask a mother-tongue speaker to prepare two texts that are written differently, each using a different version of the orthography feature you are testing. The passages must be reasonably difficult.
2. Ask individual mother-tongue speakers to read both passages aloud.
3. Listen for hesitations, repeats or errors.
4. Ask each person which text they prefer and why.
5. Write your observations after each oral test.
6. After conducting several tests, look for a pattern. If certain problems recur, do they present reading problems?

Example situation: to test the symbolization of a fricative b [β], which is contrastive with a plosive b, write one test using **bb** to represent the fricative, and another text in which the fricative is written **bh**.

Discussion: text will show things a simple list cannot. Certain features in the context may make up for ambiguities in the spelling system, so that the reader can still read with fluency and comprehension, given enough context (even a comparison of short phrases such as **two list** vs. **two lists**, **three dog** vs. **three dogs**). Rather than have someone read two different stories, you can opt for two lists of sentences, for example, written two different ways. See how much people stumble, if at all, when they come to the redundant morpheme, now unwritten.

Sarah Gudschinsky (1973:126) suggested that reading a text involves two steps:

1. recording the oral reading of a passage
2. assessing the comprehension of it

The test can then be studied as to how many real mistakes were made and why. Was the passage understood clearly? Were there any ambiguities for the reader? Did he or she have to read the passage more than once to get the meaning? An orthography which requires more than one reading is problematic.

Oral test with a wordlist

1. Ask a mother-tongue speaker to prepare two wordlists that are written differently.
2. Ask other individual mother-tongue speakers to read both wordlists out loud.
3. While the people read, listen for any hesitations, repeats, or errors.
4. Ask people which list they prefer and why.
5. Write your observations after each test.
6. Look for patterns in your data….

Example situation: the language has been previously written, with no differentiation of [b] and [β]. If the PTEST has shown complementary distribution of the two phones,

1. Ask a mother-tongue speaker to prepare one wordlist in which both [b] and [β] are written the same.
2. Ask a speaker of the language to read the words.
3. Note any problems they have. In the case of minimal pairs or other forms of ambiguity, ask the reader to tell you the *meaning* of the word/words in question.
4. Ask the mother-tongue speaker to prepare a second wordlist, in which the two phones are represented differently.
5. Ask the speaker if he or she perceives a difference in the sounds of the two. Lists of minimal pairs can be very useful but are not always possible.
6. Find out whether the speaker is consistently able to sort the words into two categories (i.e., stop and fricative).

Written tests

Dictation

1. Ask a mother-tongue speaker to prepare a wordlist or short sentences with items to test.
2. Ask a mother-tongue speaker to read the wordlist or short sentences to a small group of mother-tongue literates.
3. Have literates write the words as dictated. Note where people have a problem.

Example situation: Vowel length is conditioned before prenasalized obstruents. Mother-tongue speakers, however, hear these vowels as long. You would prefer to write only *true* vowel length in the language, because sometimes vowel length is truly contrastive, either lexically or grammatically. You give the spelling rules to a group of mother-tongue speakers and then give a dictation test to see how difficult it will be for them to write vowel length only when it is contrastive. Note trends.

To devise a multiple choice cloze test with options written out:

1. Ask a mother-tongue speaker to write several sentences containing the phenomenon in question, such as distant past tense.
2. Remove the word containing the problematic element, replacing it with a _____.
3. Write the word, spelled in the alternative mother-tongue ways you are considering, below the sentence.
4. Ask mother-tongue speakers to read the sentences, circling the spelling they prefer for each missing word.

Discussion: This type of test combines silent reading and writing. It's useful for evaluating people's comprehension of some underdifferentiated element. It is easy to assume that the context will give the reader all he needs, to compensate for that underspecification. This test may clarify his ability to *really* do this.

It also simply shows you the reader's preferences. The Tharaka language uses extra-long vowel length to indicate distant past tense. Tone is the only other indicator of that tense, but tone is not written in any of the surrounding related languages, so writing of tone is not an option. A mother-tongue speaker devises a test in which there are other cues which indicate to the reader that the event described in the sentence took place long ago. For example, ba + a + akire, '*they built*' is written two ways (**baakire**, **baaakire**), and participants are asked to write the word in a blank in the sentence, in the form they prefer. Seven or eight similar sentences are given.

Long ago, Noah and his family_____(**baakire/baaakire**) an ark.

Further testing may be required, if no satisfactory solution is found to your spelling dilemma. Try representing a morpheme such as this one, which is conveyed orally through tone as well as length, by using a punctuation mark to indicate tense/aspect at the beginning of the word (for example, * might indicate distant past tense, automatically cuing the reader to a melody and rhythm associated with this tense/aspect feature on the verb). Such

a symbol would communicate *meaning* directly to the reader, rather than pronunciation. Test again for readability (and acceptability) with several people.

Comprehension tests

   The test above is limited. It doesn't show the tester or the testees whether the distinction is necessary; it only shows what people *prefer*. To provide more insight, either for yourself or for the readers, provide two texts written with the different representations (see above) and ask comprehension questions. People can read silently or aloud. The tester gains information, either way, and the testee's awareness of comprehension issues is heightened.

Text example version 1:	Long ago, Noah and his family **baaakire** an ark…
Text example version 2:	Long ago, Noah and his family ***baakire** an ark…[1]
Text example version 3:	Long ago, Noah and his family **baakire** an ark…
Comprehension question:	When did the story take place? What word or words tell you this?

Some Notes on Tone

   Testing for tone, in particular: The real issues at stake with respect to tone and orthography are *what tonal melodies* are in contrast and *at what level of the phonology* is the mother-tongue speaker most aware of these contrasts? So discussing *how* to represent contrasts is really secondary to discussing *which* contrasts to represent and at what level of the phonology or morphology they should be represented.

> Not marking tone in any way is probably the easiest strategy for writing, but it can pose problems for reading comprehension. Whether it does pose problems for readers or not depends on how great a functional load tone bears. One way to determine how great the functional load is, is to take texts of varying degrees of difficulty, not mark tone on them in any way, and then use them to test reading fluency and comprehension. A good strategy when testing a zero marked text is to note in particular where the difficulties lie, if indeed there are difficulties. (Snider 2001:5-6)

The three things to look for when testing for any possible tone marking rules are:

1.   fluency of reading

2.   comprehension of a text (not a wordlist)

3.   spelling accuracy and ease

   It is always simplest to begin with less marking and to add whatever is needed to minimize ambiguity[2] for the reader. These are the recommended steps to take in testing.

1.   Look for ambiguity in a list of isolated words:

---

   [1]The tonal melody (marking the tense in question) is represented by a symbol preceding the verb
   [2]For more information on tonal ambiguity analysis for African languages, see Roberts, 2008.

a. Generate an exhaustive list of homographs in the language (disregarding tone at this point)

b. Look for sets of ambiguous words (homographs), and identify their respective parts of speech.

c. These sets will be your reference for all further research

2. Analyze the frequency and distribution of these homographs in texts (a variety). These are also known as tonal minimal pairs.

3. Start with a text unmarked for tone, and a strong reader who has plenty of exposure to the orthography. Perform a miscue analysis of oral reading.

a. I recommend using a tape-recorder, but testing can be documented on-the-spot as well.

b. Make sure that both your reader and listeners are completely unbiased. They don't need to know the purpose of the test, other than that we are checking to see if their writing system works well for them.

c. Use a 1st person narrative text (or some other text the reader and listeners have not seen or heard before), as clean and up-to-date as possible. Everyone will listen as a strong reader reads the text aloud.

d. Instruct the listeners raise their hands when the melody is wrong. (If the melody is wrong, also ask if the meaning is unclear).

e. On your own copy of the text, perform a miscue analysis as the reader reads aloud. Mark all mistakes, self-corrections, and long hesitations. If comprehension is questionable, the listeners and the reader can also be asked a few comprehension questions.

f. If melody irregularities are identified by the group of listeners, or if there were significant hesitations and autocorrections or confusion about meaning, use an equivalent text, which is marked for tone. Use the same reader, if possible (of course he/she must already be well-practiced with reading the tone marks).

g. Compare the results with those from reading of the unmarked text.

Evaluating Test Results

Document everything as thoroughly as possible.

If there is conflict between written results and verbal opinions expressed, give priority to written results.

If there is conflict between comprehension of a text and verbal opinions, give priority to the results of the comprehension test.

Appendix A: Determining Functional Load for Tone

Most Bantu languages will not exhibit a high functional load for tone. In Bantu A languages, and some others, however, the role of tone may be too important to be ignored orthographically. Stephen Anderson, Cameroon, uses the Bantu Phonology Tool, PTEST, to generate two lists of tone contrasts in a specific language (one list for verbs and another for nouns). These lists can be rather long, especially if the language is from Cameroon. Steve then edits them to provide a unique list of all the multiple tone contrasts. The following two paragraphs are his words (pc):

> To understand the process, imagine a simple tone triplet[1]. The computer will actually generate three tone pairs instead of a tone triplet. I correct this by editing. Also, depending on the quality of the input, the computer may generate lots of pairs that really have to do with only a single root. That root may have multiple meanings (either real homonyms or just multiple meanings of one root). I correct that as well by editing. When I am done, I have a single list of all the tone contrasts, without what I consider false data that would skew the results. Once I have my unique list of tone contrasts, then I am ready for the next step, marking on a separate sheet the number of contrasts that occur, feeling this may be a more accurate indicator of "the functional load of each contrast". In practice, this has been very helpful in deciding which tones to write and which to dare to leave unwritten (our major hurdle here).

> Sometimes a tone contrast is for more than a pair of words, so I also get rid of any homonyms or apparent homonyms so that the tonal roots occur only once in each multiple tone contrast. Once the (dual) lists of tone contrasts have been fully edited, I go through them slowly and marks each tone contrast by making a line between two tones on a rough chart, as below:

> **h**   **hl**
>
> **m**   **lh**
>
> **l**   **lx**

> Then, each time I get an additional pair of words for a contrast already indicated, I make a little slash mark on that particular line to indicate quantity of contrasts.

The first column in the table below is actual data from a 440-wordlist of Rangi nouns. Oliver Stegen (pc) shows the contrasting tone melodies for which minimal pairs have been found, and the second column gives the count ("slash marks on the line"). L=low, H=high,

---

[1]Concerning the computer writing a tone triplet as three tone pairs, 'muumba': with RM# melody it means 'Creator', with FL 'Maasai' and LL# 'barren woman'. Instead of treating the three as a triplet RM# vs FL vs LL#, the computer will list three *pairs*: 1. RM# vs FL, 2. RM# vs LL#, 3. FL vs LL#. Steve's editing triplets back into the contrastive tone melody list will lengthen the chart. The benefit: the linguist could see three-way distinctions immediately at a glance, which would be lost when displaying tone ***pairs*** only.

M=mid, F=falling, R=rising tone and # =a boundary tone phenomenon (e.g., 'x' in Anderson's example above).

FL# vs HM#	| = 1 instance
FL# vs LL#	||||| = 5 instances
HM# vs LL#	|| = 2 instances
HM# vs RM#	| = 1 instance
RM# vs LL#	| = 1 instance
FL vs HH	| = 1 instance
FL vs LH	|||| = 4 instances
FL vs RH	| = 1 instance
HH vs LL	| = 1 instance
HH vs LH	| = 1 instance
HH vs RH	| = 1 instance
RH vs LH	| = 1 instance

It can now quickly be seen that in Rangi, falling versus low tone melodies may need to be distinguished lexically (high instance of **FL#** vs **LL#**). By contrast, even though both **LL#** and **LH** occur as distinct tone melodies in Rangi, they *never* contrast. Consequently, they can be grouped together and marked with the same pattern (in the case of Rangi, no tone mark at all versus tone mark on the first stem vowel for falling melodies). Similarly, **HH** vs **HM#** never contrast, nor do **RH** vs **RM#**. So basically, this chart shows us that the functional load of tone is a lot higher on the first stem syllable (contrasting **L**, **H**, **F** and **R**) than on the second/final (where no consistent contrast could be identified, and consequently, tone is never marked on the final stem syllable in Rangi). A revised chart, taking into account only the tone on the first stem syllable, would then look like this:

F vs H	|| = 2 instances
F vs L	||||||||| = 9 instances
F vs R	| = 1 instance
H vs L	|||| = 4 instances
H vs R	|| = 2 instances
R vs L	|| = 2 instances

At the end of this process, the following can be seen:

1. Which TONE PAIRS actually contrast, which do not

2. Which TONE PAIRS have a high functional load lexically, and which have lower functional loads

The result of this process: clues as to which tones need to be distinguished from other tones and which do not. "Depending on the language, one can often lump several tones together as carrying minimal functional load. Sometimes it is good to leave the largest *group* of tones unmarked in order to minimize tone marking" (Stephen Anderson, pc).

The dual lists function completely independently. Verbal and nominal tone usually involve completely different issues. One possible outcome, then, of the tonal contrast counts may be a decision to mark tone for some verbal forms or for verbal roots, while not marking it for nouns.

The final step involves going through the minimal pairs of words which one has decided not to mark differently (orthographic homonyms), to see if the two meanings are likely to be confused in context. If there is little chance of confusing the orthographic homonymns in textual material, then one feels better about minimal marking of tone. If certain words are potentially confusing, however, it is better to find some creative way to differentiate them orthographically.

This is often the case with grammatical words, and small words which carry less segmental or context-dependent information. They often occur frequently in text and affect people's oral and silent reading fluency. This is why, for example, it is so important in the Spanish language for the stress difference between <esta> 'this' and <está> 'it is' to be reflected in the orthography.

Example of ambiguity created from failure to provide tonal information in a language with heavy functional load for grammatical tone

Lendu (Central-Sudanic, DRC)

ma bbi bbi	'I walked'
ma bbi bbi	'we walked'
ma bbi bbi	'I will walk'
ma bbi bbi	'we will walk'
ma bbi bbi	'I should walk'
ma bbi bbi	'we should walk'
ma bbi bbi	'I am walking'
ma bbi bbi	'we are walking'

Appendix B: Glossary of Linguistic Terminology

Clitic: a monosyllabic (in Bantu languages) word which is phonologically bound to another whole word.

Compensatory lengthening: a phonological phenomenon involving the deletion of one segment accompanied by an increase in the length of another, usually adjacent to it, thus preserving syllable length or "weight". This happens often with vowels in Bantu languages, in which vowel length, while not lexically contrastive in a certain instance, is phonetically created. Automatic vowel lengthening in the penultimate syllable is one example; another occurs when a consonant is elided between two vowels. A third source of compensatory lengthening takes place before a prenasalized stop. Fourth, it can take place after a glide formed from a vowel (Gardner: 2005:17).

Copula: a linking verb which has little independent meaning. It serves to relate different elements of the clause structure, especially subject and complement. A prime English example is the word "is", though there are many others which function similarly.

Dahl's Law: a rule of consonant dissimilation. A voiceless stop becomes voiced if the consonant in the next syllable is voiceless. (Example: Gikuyu, from Kikuyu). In other words, "When two successive syllables in a stem each begin with a voiceless plosive, then the first of these becomes voiced" (Schadeberg 1999:391). This process usually occurs in prefixes, as in the Kikuyu example above, in which the prefix /ki-/ becomes /gi-/ before /-kuyu/. In some languages only traces of this law remain; in others it has never applied.

Functional load: relative importance of a linguistic feature in conveying grammatical or lexical information. Example: If tone is the only indicator of passive tense for Kitharaka verbs, its functional load is very high in that context. If both tone and vowel length are indicators of passive tense, then tone's functional load is lower. Lendu, a Central-Sudanic language, provides an example of grammatical tone carrying a very high functional load. This is what the Lendu orthography would communicate, without tone depicted in the writing system. See also BOM Functional load.doc for a method to determine functional load of tone for Bantu verbs.

    <**ma bbi bbi**>   *"I walked"*

    <**ma bbi bbi**>   *"We walked"*

    <**ma bbi bbi**>   *"I will walk"*

    <**ma bbi bbi**>   *"We will walk"*

    <**ma bbi bbi**>   *" I should walk"*

    <**ma bbi bbi**>   *"We should walk"*

<ma bbi bbi>   *" I am walking"*

<ma bbi bbi>   *"We are walking"*

Grammaticalization: the expression of a semantic contrast using grammatical forms rather than content words. Grammaticalization takes place over time in a language, and words which carry distinct meaning, such as nouns and verbs, can erode and become functors.

Hiatus: the occurrence of two vowel sounds without pause or intervening consonantal sound

Katupha's Law: "Given two consonant-groups in a word, separated by a vowel and themselves aspirated, and provided that they are in the same root, then one (and normally the first) is deprived of its breath feature," i.e. it is de-aspirated (Schadeberg 1999:391).

LWC: language of wider communication, such as French, Swahili, Portuguese or Luganda

Meinhof's Law (also known as Ganda Law): prescribes the formation of geminate (long) nasals when a nasal adjoins a consonant, and the assimilation of the nasals to the point of articulation of the underlying consonant. This happens when post-nasal voiced consonants become nasalized, e.g. Ganda E15(J) (Hyman 2003:51):

/n≠bomb-a/→ m≠momb-a   *'I escape'*

/n≠limb-a/ → n≠nimb-a   *'I lie'*

/n≠jung-a/ → ɲ≠ɲung-a   *'I join'*

/n≠gend-a/ → ŋ≠ŋend-a   *'I go'*

There are several variants of this rule, one of which (the Kwanyama Rule) removes prenasalization from the second consonant in the root (Schadeberg 2003:148).

Morphophonemic alternation: An alternation between phonemes when two morphemes are adjoined

Proclitic: a clitic which depends upon a following word (i.e., an article such as the, a, or an in English)

Spirantization: process in which a language turns obstruents followed by *i* and *u* into strident fricatives or affricates

Vowel harmony: a common phenomenon in Bantu languages in which vowels are divided into two mutually exclusive groups. The vowels of one group tend to occur with each other in words, to the exclusion of vowels of the other group. Vowel harmony can be seen as the spreading of a phonetic feature (back, high, round, advanced tongue root) to certain vowels, and not to others, within a word. The spread is usually leftward from the root.

Appendix C: Morphology vs. Phonological Rules

You can have your cake and eat it, too: An orthography that meets the needs of both mature and immature readers

Keith Snider
SIL International
Canada Institute of Linguistics
(Lecture given to Orthography class, SIL-UND, June 29, 2005)

1.   Introduction[1]

Mature readers need a constant visual word image

Mature readers read by sight (Venezky 1970), and do not take the time to sound out the words they read. For this reason, a good orthography maintains a constant word-image. This helps minimize the effort it takes for a developing reader to memorize the shape of each word.

Immature readers need to spell words the way they perceive them to sound

Beginning readers often sound out their words. For this reason, a good orthography bases the spelling of its words as closely as possible on the way the native speaker *perceives* the words to sound.

2.   The problem[2]

There can be tension between the needs of mature readers and the needs of immature readers.

If we write words the way they are pronounced in order to help immature readers as much as possible, this can cause problems for mature readers. This is because the pronunciation of words changes in different environments and writing words with these changes means that the mature reader will not have a constant word image. Here is an example from Chumburung, a language spoken in Ghana.

Words spoken in isolation

| kofi | 'Kofi (person's name)' |
| kuŋu | 'head' |

---

[1]This paper is an expansion of §2 in Snider (2001). Since I am the author of both works, I have taken the liberty of including in the present work a number of sections verbatim from the earlier work. I have also taken the liberty of editing some of these sections liberally to better fit the present purposes. This has made it difficult to acknowledge which material comes from the former work and which material is new and so I have not done that.

[2]This is true only for alphabetic orthographies. A logographic orthography, in which there is a one-to-one correspondence between a symbol and a word or morpheme, sidesteps this issue, of course.

Words spoken in context

    kofu kuŋu   'Kofi's head'

Notice that the [i] in 'Kofi' changes to [u] in 'Kofi's head.' One thing that nonlinguists are not always aware of is that there is often a difference between the way a native speaker *pronounces* a word and the way he/she (hereafter he) *perceives* it to sound. In the case of this Chumburung example, while the native speaker pronounces 'Kofi' with an [i] in isolation (i.e., citation form) and with a [u] when it comes before the word for 'head,' he doesn't realize he is making this change. In other words, he does not perceive this change.

The "trick" in developing a good orthography is to write words the way the native speaker *perceives* the word to sound, not necessarily the way the native speaker actually pronounces them. This is great for the immature reader. There is an added bonus, however, and that is that when we write this way, we are also able to maintain a constant word-image. Writing the way native speakers perceive the language to sound meets the needs of both mature and immature readers with the same orthography.

Developing an orthography would be relatively easy if the pronunciation of the sounds in the language was stable. Unfortunately, as we see in the Chumburung example above, this is not the case. Sometimes, as in the Chumburung example just discussed, the changes are below the level of native awareness (i.e., the native speaker isn't aware that he is saying the words differently). Here is another example of a change that the native speaker isn't aware of, this time some plural forms from English.

Voiceless	Voiced
cap–s	cab–z
bit–s	bid–z
buck–s	bug–z

Here we see that the native speaker of English pronounces the plural form as (voiceless) *s* when it follows voiceless consonants and as (voiced) *z* when it follows voiced consonants. In other environments, English has a contrast between *s* and *z*, as in *sip* and *zip*. However, this contrast is lost when the sounds immediately follow a consonant and which of the two sounds is produced is totally predictable from the environment. Although native speakers are aware of the difference between *s* and *z*, they are usually not aware (unless it is pointed out to them) of the differences in the above words.

At other times, however, the native speaker is well aware that he is saying sounds differently. A good example of this occurs in the English negative prefix *in-*.

    im–possible

    in–tolerant

    il–logical

    ir–reverent

In these examples, the native English speaker is well aware that he is saying an *m* before *possible* and an *r* before *reverent*, etc. He is so aware of this difference in fact, that he probably doesn't even realize that these variants are all forms of the same prefix, but rather

he thinks of them as different prefixes. So, some rules produce differences that the native speaker perceives (we'll call these *word rules*), and other rules produce changes that the native speaker does not perceive (we'll call these *phrase rules*). Since this is the case, in order for an orthography to represent only sounds that the native speaker perceives, we need to be able to distinguish between these two types of rules.

3. Data

In this section, there are two data sets of English examples. The first set demonstrates a rule called English Flapping. This is a phrase rule and it produces changes that the native speaker of English does not perceive. The second set demonstrates a rule called English Trisyllabic Laxing (hereafter TSL). This rule is a word rule and it produces changes that the native speaker of English does perceive.

3.1 Changes the native speaker does not perceive (phrase rule)

English Flapping (Kenstowicz, 1994:195)

a) aɾom   cf. atom

b) meeɾ-ing   cf. meet

c) whaɾ is wrong? cf. What

Here are the relevant facts concerning English Flapping:

a) An intervocalic coronal stop changes to a sonorant coronal flap *r* when it occurs between two vowels, the first of which is stressed.

b) While /t/ is a contrastive sound (phoneme) in the language, [ɾ] is not.

c) Native speakers do not realize that they are saying [ɾ] differently from [t].

3.2 Changes the native speaker perceives (word rule)

English Trisyllabic Laxing (Kenstowicz, 1994:196–197)

a)	divine	[aj]	divin-ity	[ɪ]
	serene	[ij]	seren-ity	[ɛ]
	profane	[ej]	profan-ity	[æ]
b)	vile	[aj]	vil-ify	[ɪ]
	clear	[ij]	clar-ify	[ɛ]
c)	rite	[aj]	rit-ual	[ɪ]
	grade	[ej]	grad-ual	[æ]
d)	tyrant	[aj]	tyrann-ous	[ɪ]
	penal	[ij]	penal-ize	[æ]
e)	tyrant	[aj]	tyrann-ous	[ɪ]
	fable	[ej]	fabul-ous	[æ]

Here are the relevant facts concerning English Trisyllabic Laxing (TSL):

a) The rule shortens and laxes a long vowel when it is followed by two syllables, the first of which is unaccented.

b) not adequate. Forms like *nightingale, stevedore,* and *ivory* do not shorten the ini tial vowel. The change is only conditioned by the addition of a suffix.

c) Not all suffixes trigger TSL (cf., *brave/brav-ery, might/might-ily, pirate/pirat-ing*).

d) Within the class of suffixes that do trigger TSL, there are still idiosyncratic lexi cal exceptions (e.g., *obese/obes-ity*).

e) Native speakers are aware of these changes.

Trying to determine which sounds a native speaker perceives and which sounds a native speaker does not perceive can be extremely difficult since the matter can be very subjective. For this reason, I am including questions below that one can use to determine more objectively which is which.

4. How to discover word rules [3]

In order to help discover word rules, that is, rules that produce sounds that the native speaker perceives, ask yourself the following questions:

1. Does a given rule apply across the board without exception, or are there lexical exceptions? In other words, in the same grammatical and phonological environment do some words undergo a particular rule while others do not? If there are lexical exceptions to a particular rule, then it is important to spell words that undergo that rule the way they sound *after* the rule has applied. In the examples above of English TSL, not all suffixes trigger TSL (cf., *brave/brav-ery, might/might-ily, pirate/pirat-ing*). This means that there are whole classes of lexical exceptions. Then within the class of suffixes that do trigger TSL, there are still idiosyncratic lexical exceptions (e.g., *obese/obes-ity*). This means that the TSL rule produces changes that the native speaker perceives.

2. Does a given rule lack phonetic motivation? For most rules, we can see a phonetic reason for why the rule's output sounds the way it does. For example, sounds often assimilate to certain qualities of other sounds. If however, a particular phonological rule does not have any phonetic motivation whatsoever, then spell the word the way it sounds *after* the rule has applied. Here is another example from English.

   electrik   electris–ity

   In this example, the final *k* of 'electric' changes to *s* when the suffix –*ity* is added. There is really nothing in the phonological environment that would induce a *k* to change to *s* before –*ity*, so it is obvious that the rule lacks phonetic

[3]This section and the next owe a great deal to the theory of lexical phonology. For a detailed overview of lexical phonology, the interested reader is referred to Kenstowicz (1994). For a clear presentation of the differences between rules that produce sounds that the native speaker perceives (viz. lexical rules) and rules that produce sounds that the native speaker does not perceive (viz. postlexical rules), see Pulleyblank (1986).

motivation. This is a tip-off that the native speaker is aware of the change that the rule makes. The implication for orthography is that we should spell the two words as follows: *elektrik* and *elektrisity*. Notice that this does *not* permit one to maintain a constant morpheme-image. It does, however, permit one to maintain a constant word-image, which is what is important to mature readers. In this case, the word *elekrik* will have a constant word-image, and the word *elekrisity* will also have a constant word-image.

3. When a given rule applies, does it only apply when a prefix or suffix is added? In other words, does the rule have to apply across a *morpheme* boundary? Note that we are NOT talking about a word boundary. If the rule must apply across a morpheme boundary, then the word should be spelled the way it sounds *after* the rule has applied. Again looking at the English TSL example, we see that specification of the environment of the change in purely phonological terms is not adequate. Forms like *nightingale, stevedore,* and *ivory* do not shorten the initial vowel. The change is only conditioned by the addition of a suffix. Since the rule is conditioned by a suffix and would not otherwise apply, this tells us that the native speaker is aware of the change that the rule produces. hence we should spell these words with the changes.

5. How to discover phrase rules

In order to help discover phrase rules, that is, rules that produce sounds that the native speaker does not perceive, ask the following questions:

1. When a given rule applies, is the new sound it produces one of the contrastive sounds in the language? If the output of the rule is not a phoneme, then spell the word the way it sounds *before* the rule has applied. In the English Flapping rule above, the sound *ɾ* is not a phoneme in the language and so it should not be written with the change because native speakers are not aware of the noncontrastive sounds in their language.

2. When a given rule has applied, do native speakers think that the sound that results is the same as or different from the sound that underwent the rule. If they think it is the same sound, that is, they don't realize that anything has changed without having the change pointed out to them, then spell the word the way it sounds *before* the rule has applied. Again in the case of the English flapping rule, the native speaker is not aware that he is not saying [t] in these forms. This is another reason why we should write the form the way the native speaker perceives it to sound (i.e., *t*) as opposed to the way he actually pronounces it (viz. *ɾ*).

3. When a given rule applies, is it necessary to refer to the internal structure of the phrase in the rule's environment? In other words, does the rule apply across a *word* boundary? Note that we are NOT talking about a *morpheme* boundary. If the rule applies across a word boundary, then the word should be spelled the way it sounds *before* the rule has applied. This means that the word-image does not change even though the word may be pronounced differently in the context of other words. For an example of this, recall the Chumburung example above, repeated here for convenience sake.

Words spoken in isolation

   kofi   'Kofi (person's name)'

   kuŋu   'head'

Words spoken in context

   kofu kuŋu   'Kofi's head'

The rule that changes /i/ to [u] applies across a word boundary. This tells us that the native speaker is not aware of the change and so we should not write the change. This permits the word *kofi* to have a constant word-image and not be realized in two different ways.

Following the above criteria helps to establish where word breaks should occur[4] and eliminates the representation of low-level allophonic variation. It also eliminates the effect of rules that apply across word boundaries and the effect of phrase initial and phrase final phenomena. This allows one to: a) maintain a constant word-image and thereby meet the needs of mature readers, and b) write words the way the native speaker perceives them to sound and thereby meet the needs of beginning readers.

A word of caution. Although the above principles present a way to rigorously determine what should be written and what should not be written, it offers only a promising beginning point for people to try. As with everything else in orthography design, try something and then test it thoroughly.

References

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell Publishers.

Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: D. Reidel Publishing Company.

Snider, Keith. 2001. Linguistic factors in orthography design. In Ngessimo M. Mutaka and Sammy B. Chumbow (eds*.), Research material in African linguistics: Focus on Cameroon*, 323–332. Köln: Rüdiger Köppe Verlag.

Venezky, Richard L. 1970. Principles for the design of practical writing systems. *Anthropological Linguistics* 12:256–270.

---

[4]Only those criteria which do not make crucial reference to word boundaries are useful in determining where word breaks occur. Otherwise, this is a circular argument.

Part 2

A Participatory Approach to Orthography Development

Foreword

This procedural guide for linguistic analysis describes a strategy for orthography development using a workshop approach. It is very practical in nature. Part 1, the Bantu Orthography Manual, should be used along with it. This procedural guide has a target audience: the "Ordinary Working Linguist." Chapter 1 explains the rationale of Kutsch-Lojenga's participatory approach to orthography development, and gives the big picture. Chapters 2–4 provide step-by-step instructions for carrying out a participatory approach via a series of workshops.

Acknowledgements

Native speaker insight is essential to the success of the participatory approach, which has a spiraling,[1] informal curriculum. Native speakers are introduced to features of the languages in their area before linguistic analysis begins, and as/after discoveries are made. They are helped to see the impact of spelling choices upon writers and readers, ongoingly. Optional lectures and templates are provided in appendices.

---

[1] A curriculum which is additive and repetitive, constantly building upon what people learn, and revisiting those concepts with greater depth is known as "spiraling."

Chapter 1: Overview of Procedures

The goal

Develop an orthography for a Bantu language group, in partnership with speakers of that language (laying a linguistic and educational foundation for the application of that orthography to reading and writing).

Foundational Assumptions for Orthography Development

- Community ownership/involvement in every stage of the orthography development process

- The centrality of linguistic analysis to orthography development, inclusive of phonology, morphophonemics, grammar and discourse (since these aspects of a spoken language overlap and influence one another on the surface, they will seem to compete with one another for prominence in a writing system)

- Speakers' perception should play a significant role in orthography decisions. That perception can be developed and enriched for those who take part in the orthography development process

- Orthography-in-use as goal, and also as means for constant feedback and evaluation

- Questions of readability and writeability will be considered throughout the development process

- Revisions will be ongoing, and will necessarily reflect the political, educational and social context of the writing system

Goals

The approach outlined below was developed by Constance Kutsch-Lojenga. Details and suggestions have been added by Leila Schroeder and Karen van Otterloo. The general goal of Kutsch-Lojenga's approach is to allow participants to discover for themselves the phonemes of their language, by systematically sorting words according to the various sounds found in them. We have incorporated the goals of alphabet development and documentation of linguistic discoveries into the process described below.

Each topic is introduced generally, followed by detailed procedural instructions. Each linguistic feature studied is captioned, i.e., Vowels, Week 1. Below each linguistic feature caption are the instructions, including documentation of linguistic discoveries in Workbook Form, i.e., Fill in Table 2, the Vowel Combinations Chart

Linguistic rationale

Since there is a lot of variation between Bantu languages as to the rules which produce surface forms of morphemes, the participatory approach to orthography development will

give significant attention to morphological and morphophonological issues before writing rules are determined.

This procedure was originally developed in East Africa, in the Swahili-speaking area, and therefore the document contains many references to Swahili as the Language of Wider Communication. However, the procedure can be adapted to other parts of the Bantu-speaking world, where other languages function as the LWC).

Overview

The following section outlines a methodology for developing an orthography for a Bantu language. Whether the focus of study is one language or a cluster of languages, the approach will be the same.

Initial Orthography Development: studying phonology and morphophonology

Discovery of contrastive sounds

vowels

consonants

tones

Phonotactics: the combinations and combinatory restrictions of Consonants and Vowels, particularly in root-initial position.

This study is needed

1. for the discovery of cases of complementary distribution:
   - in terms of CV-combinations
   - in terms of position of a C in the root structure
2. as necessary background material for the creation of exercises in primers

Morphology and Morphophonology

After establishing its phonemic inventory, the linguist wants to discover the lexical, or word-internal, morphophonology of a language. Two aspects of morphology require study: inflection and derivation. For Bantu languages, the study of morphophonology, both at morpheme boundaries within words and in clitic constructions, the main example of which is the associative construction, is also important.

This information is needed so that sound changes in speech which are caused by phonological processes can be reflected well in writing. Most of the results of these word-internal phonological and tonological processes will be written according to surface forms. The phonological changes which each affix produces and/or undergoes needs to be written up in an orthography guide (or "writers' guide"). Once in a while, such phonological processes produce surface sounds which are not found in the inventory of underlying contrastive sounds. These need very specific consideration with respect to orthography. Sometimes – especially

in the associative and other clitic constructions – questions on word boundaries need to be taken into consideration together with the morphophonological sound changes.

Fortunately, Bantu languages are quite transparent in their morphological and morphophonological structure, though there is great variation in the ways they handle them.

Approach: Participatory Research

The people participating in this research should be made aware of every aspect of the findings immediately, as the process of research is going on. They should participate and give their opinions and insights, some of which will help the research move faster, and promote learning for all. The approach reflects a "learning by doing" philosophy.

Linguistic analysis for this approach begins with word collection, using native intuition for initial spelling. This is followed by electronic data entry, and then moves to printing and sorting of word cards based upon phonotactics and target-language speaker perception. Alphabets are developed and modified as necessary. Finally, participants discover their grammar together and, guided by Chapter 3 of Part 1 (Bantu Orthography Manual), devise spelling rules which try to encompass their language's morphology as well as its sound system.

Caveat: Native speaker participation with intuitive writing as the basis for the original wordlist transcription *does not* imply that native speakers are asked to make suggestions as to how they want to write the more linguistically and pedagogically complex phenomena. The Manual provides a narrow list of grapheme and word break choices, and describes the issues for each. Participants should, after a good presentation and discussion of the issues involved, choose between the *few* options offered them. Testing of readability as well as acceptability will follow such decisions if questions remain.

Chapter 2: Data Collection

Each topic is introduced generally, followed by detailed procedural instructions. Each linguistic feature studied is captioned, i.e., Vowels, Week 1. Below each linguistic feature caption are the instructions, including documentation of linguistic discoveries in Workbook Form, for example, "Fill in Table 2, the Vowel Combinations Chart."

Word Collection

Goals of word collection phase

Collect around 1,700 words, nouns, and verbs, as written by speakers of the language. These will be used in the initial determination of symbols needed in the alphabet, as well as being a resource for primer making and other language-development tasks. Speakers of the language will represent the sounds of their language using the symbols of their language of wider communication (LWC). By looking at inconsistencies in their transcription and at underdifferentiation (using a single symbol to represent two different and significant sounds), we and they gain a good idea of the areas which will require study. Target-language speakers are likely to underdifferentiate sounds—especially vowel sounds—consistently. Native speaker intuition can be better than IPA transcription when orthography is the ultimate goal, because the linguist will not know which phonetic differences are significant, nor which sounds are the same to the target-language speaker. His/her phonetic transcription will also yield a less than 100% accurate transcription, but it will normally be inaccurate in a consistent manner.

The following description of the process has been gleaned from a detailed report written by Susanne Krüger, based upon the first series of word collection workshops in Mara region of Tanzania in 2006.

Constance Kutsch-Lojenga's original participatory approach did not include wordlist collection via computers. It was simpler, with two goals in mind: native speaker insight and intuitive orthography development accompanied by linguistic analysis. We have described the use of Toolbox,[1] a language software tool, for the word collection and documentation process, in an effort to gain the beginnings of a dictionary for each language, as a resource for primer development, and as a means of preserving spelling decisions. Toolbox and FieldWorks, a similar software package, are able to give input directly to the Bantu Phonology Tool. A software program other than Toolbox will require some procedural adjustments.

Word collection: wordlists vs. semantic domains

Word collection will be done using the Africa Area 1,700-item wordlist, in Swahili. Pitfalls of reliance upon a wordlist: many times the lexicon of the target is not an exact match for the LWC gloss. People will try to make a fit, and the result can be less-commonly used loanwords, phrases, ambiguities of various sorts, and lots of blanks.

---

[1]Field Linguist's Toolbox is software that provides integrated data management and analysis for the field linguist, developed by SIL International to aid in linguistic data collection and publishing.

Due to this challenge, an introductory activity using a semantic domain approach may encourage participants to think more freely. For instance, if they have three ways to express the Swahili gloss in their mother-tongue, they may offer more interesting mother-tongue vocabulary for the list (and less Swahili), whether or not it matches the Swahili gloss exactly.[2] Try this with noun categories, such as "furnishings," "hunting tools," "musical instruments," to get people thinking of their own rich vocabulary.

Whether the collection of vocabulary is done by wordlists or by semantic-domain study, the speakers of the language should write the words first of all themselves. This emphasizes local ownership of the project from the start and also helps the linguist because people will make systematic mistakes, rather than the random ones outside researchers would make.

The specific vocabulary desired for Bantu wordlists: nouns in their singular and plural forms (so as to be able to determine the root), and verbs, in their infinitive form. Ignore tone at this point.

Once the first collection of vocabulary has taken place, one should weed out compound nouns. Their phonological and morphological structure may exhibit irregular patterns. They should therefore be studied later, in light of the inventory of sounds established.

All vocabulary collected should be divided into NOUNS, VERBS, and OTHER. In phonological research, one first needs to focus on the Noun Roots and the Verb Roots, i.e., monomorphemic units. It is important to correctly identify the word form of the collected lexical items in the target languages especially when using a wordlist. The word form of the translated lexical item might not be the same as the gloss word used for prompting (especially if the wordlist used is in a language from a different language family from the target language.)

With Bantu languages, the citation form of nouns consists of a root with a prefix. Basic morphological study is needed at the onset, to determine the root structures, which will form the basis of the phonological research. Early phonological analysis should take into consideration the transparent forms of the common canonical root structures.

> NOUNS: -CVCV (and -VCV and -CV will be treated later)

> VERBS: -CVC-a (and -VC-a and -CV-a or -C-a later)

In addition to weeding out compound nouns, set aside derived forms, particularly longer verbs with derivational extensions and also agentive nouns, often formed with the suffix-i, which may cause morphophonological changes in the root. These forms can be entered in the electronic data recording system, but should not be printed out onto cards for sorting at the Orthography I Workshop.

Roles of Participants

A local community can appoint an MC for the word collection workshop (in Tanzania it is helpful to have participants elect a chairman and secretary at the beginning of the workshop). This person calls the meeting to order, organizes welcomes, devotions, and closing

---

[2]Kutsch-Lojenga actually suggests using semantic domains for noun elicitation, and wordlists for verbs ("Working document for Phonological Research in BANTU languages," presented at Bantu Orthography Working Group in Dallas, November 8, 2005).

remarks, introduces people, and generally leads through each day. This role can also be taken by the word collection team leader.

The word collection team, charged with word collection, can perform the following roles and duties:

Instructor

The instructor is responsible to explain the workshop format to the group. He/she explains the methodology of the word collection to the participants and helps organize the groups. This person is also responsible for administering a short Swahili spelling test at the beginning of the workshop.

This job can be done by the team leader. A very fluent, natural-sounding speaker of the language of wider communication is valuable at this point, both for explaining the process of word collection and for good pronunciation of the dictation test words.

Computer Scribe

All collected words are typed into Toolbox in order to store them and print out cards for the orthography workshop. The computer scribe is responsible for entering the data, backing it up every day, and printing out the cards after weeding out all doubles. This process is time-consuming and the scribe will need the help of others at times.

Contact person with hosts

This person is responsible to liaise with the hosts or the chairman regarding meals, lodging, time frame, and any other needs or problems that arise. This is an important and demanding role, taking up a considerable amount of time for the person involved. It should be a fulltime role for someone on the team. If four groups are anticipated, meaning that four facilitators are needed, then a fifth person is needed primarily to work with the hosts as contact person, and possibly to fill in for the computer scribe or facilitator.

When the team is pressed for time, participants may be divided into three groups instead of four. During one word-collection workshop only three groups were formed, and the team was still able to finish work in good time and spend more social time with the host and the participants, building relationships.

Group scribe

Each group needs at least one scribe, a speaker of the target language who writes down the collected words on a blackboard (so all target-language speakers can see and give input) in an orthography as close to Swahili (or another language of wider communication) as possible.

Scribes may initially be chosen according to their performance in a spelling test given via the language of wider communication. However, a Swahili spelling test, for example, is not always a good indicator of a good mother-tongue speller. Several times younger people have been chosen as scribes because of their faultless performance in the spelling test, but when called on to spell mother-tongue words, they may take a long time basically needing to get the words spelled out for them by others. In these cases, the person suggesting the spellings can be asked to take over after some time (to give the first spellers a chance to get

used to the job and in order to see if, with some time performance, they would get better). Generally, the best spellers (and writers) "rise to the surface" within a few hours.

Discussion facilitator

Another role that needs to be taken care of in each group is the role of discussion facilitator. This person needs to encourage the group to think creatively and "out of the Swahili box" in order to get as many usable words as possible. The wordlist should not restrict the writing down of related words. The facilitator should also make sure that all members of the group have opportunities to voice their ideas and opinions.

This role was held by the team members. Initially, the idea was that in the course of the workshop this role could be taken on by a member of the group. In some instances this happened naturally. However, the team members must keep an eye on the proceedings the whole time. It is not advisable for a group to keep working without the presence of a word-collection team member.

The groups

The preferred group size is between 3 and 5 participants (as specified by Constance Kutsch-Lojenga, personal communication).

Recommended Schedule

The following daily schedule is proposed for each workshop:

First Day:

10.00 am	Registration, Welcome, Chai
11.00 am	Introductions, Devotions
11.30 pm	Introduction to workshop
12.00 pm	Introduction to word collection, trial run, spelling test
1.00 pm	Lunch
2.30 pm	Work in groups
4.00 pm	Closing prayer
7.00 pm	Dinner

Following Days:

8.30 am	Devotions
9.00 am	Word collection in groups
10.45 am	Chai break
11.15 am	Word collection in groups
12.45 pm	Prayer time
1.00 pm	Lunch
2.30 pm	Word collection in groups

4.00 pm	Closing prayer
6 or 7.00 pm	Dinner for those not going home at night

The word-collection team meets every day after closing for debrief, typing in of words and preparation for next day.

Sessions longer than about 1.5 hours are very hard for the participants and some breaking up of the work is helpful. The early finish each day allows those participants who stay at home throughout the workshop to reach home in good time and allows the team to type in words and prepare for the next day.

It should be possible to elicit 2,000 words in a four-day workshop using three groups.

Elicitation process

Word-collection teams should begin with a semantic domain approach before attempting to collect words via the wordlist. People find it very hard to change from translating words from a list to freely associating words. The wordlists, on the other hand, give more structure to the process, a sense of progress and completion.

When the wordlist process begins, each group is given part of the BD Standard 1,700-item wordlist (a current soft copy is available upon request from the Bantu Department, SIL). The group facilitator reads the Swahili word and the group comes up with a MT translation and writes this down (on the blackboard). The facilitator then copies the word exactly from the blackboard onto the blank wordlist. Participants may give more than one word if there are multiple ways of translating the LWC word.

Information noted for every word is:

- Gloss in MT (Sg/Pl for nouns, Infinitive for verbs)
- Swahili (or LWC) gloss
- Noun, verb, other

At the end of each workshop a "clean-up session" is held. During this session, a member of the group (not the scribe) reads the mother-tongue words out loud to the whole group and the group gives a translation into Swahili. During this time all the necessary information on spelling, word break and elision or hyphenation have been noted. Corrections supported by the whole group are marked and the final version typed into Toolbox. It is possible to use a printer (powered by the car through an adapter), printing only the *mother-tongue glosses* from Toolbox. This list is easier to read than a handwritten one, and a printout of the LWC gloss is not helpful or necessary.

Any observations the word-collection team can make during the word collection about correspondence between phonetics and the symbols used by the transcribers will be helpful. The team should be especially observant during the reread—when reading their own writing, are people struggling with issues of vowel length, vowel quality, tone, etc? How often do they read a word with correct segments but the wrong tone, wrong vowel quality, or wrong vowel length? What, if any, seems to be the "default" tone pattern people use when guessing at the pronunciation of a word they do not recognize?

Data Entry

Computer tools to assist in word collection and card printout

In order to facilitate the collection of the 1,700-item wordlist and the subsequent creation of a sorting card for each word, a Toolbox database and project have been created together with wordlist documents.

Printed word collection forms

During the word collection workshop, speakers of a language can record their words on a printed wordlist form. A document of the 1,700-item wordlist with Swahili is available (1700 swa-blank 0701).

Keyboarding words

Following the transcription of the words on the printed wordlist forms, computer scribes can record the words in the Toolbox database. In the Toolbox Word Collection project, there is a database of the 1,700-item wordlist. Each record contains various fields already filled in (English gloss, Swahili gloss, wordlist index number, and semantic domain index number). The lexeme field is blank and this is where the vernacular word is keyboarded.

Before beginning data entry, one needs to create an "empty" wordlist database for each language. In the project, select the master wordlist database "Master Wordlist.db." Select the filter "00 1700 List." Then go to File/Export and select "Blank Wordlist Bantu lg." The filter should be set to the current database filter. Create a database for each language being processed and change the filename appropriately.

At the end of a session, when the vernacular data have been verified and transcribed, the Toolbox database can be filled in. If the database is sorted by the wordlist index number (\ is-1700), the user can proceed through the wordlist as it is found on the printed handouts.

For noun words, the singular is placed in the lexeme (\lx) field and the plural form in the plural field (\pl). If there is only a plural form, this should be placed in the lexeme field and the plural marker removed.

For verbs, the "root" form should be place in the lexeme field and the infinitive form placed in the vernacular paradigm field (\pdv) following the infinitive paradigm label (\pdl inf) field. An option is to place the infinitive in the lexeme field and ignore the paradigm fields.

If the data in a record does not reflect the language, it can be edited. For example, if the Swahili gloss needs to be changed or modified, the text in the Swahili gloss field (\gswa OR \gr) can be edited.

For each record that will be useable in the orthography sorting process (nouns and verbs only), the Card (\card) field should contain a "yes." If the word is not useable and does not need to be printed on a card, the Card field should contain a "no."

If multiple people are entering data, each can have a separate database and the databases can be merged at the end of the data entry session. The updated database can then be distributed to each computer scribe (for Toolbox v.1.5.1 and above).

Printing cards

When all the data has been keyboarded, the sorting cards can be printed. There are filters in the Word Collection project for Nouns, Verbs, and Other. A Word document is created. This can be reviewed and edited as needed before printing the cards.

To print Noun cards:

1. Set the filter to "00 ps nouns Cards"

2. Go to File/Export, choose "wordlist cards" export name. The filter should be set to "Current Window's records."

To print Verb cards:

1. Set the filter to "00 ps verbs Cards"

2. Go to File/Export, choose "wordlist cards" export name. The filter should be set to "Current Window's records."

To print other cards:

1. Set the filter to "00 ps other Cards"

2. Go to File/Export, choose "wordlist cards" export name. The filter should be set to "Current Window's records."

Other printouts

Other printouts can be created in the Toolbox project.

File/export/Multi-Dictionary Formatter	Standard dictionary type printout
Browse view	List of fields

The following can be created from the Master Wordlist.db:

File/export/List English-Swahili (RTF)	List with English and Swahili
File/export/List Swahili-blank (RTF)	List with Swahili and (blank)

Challenges:

A limited number of word roots in Bantu languages. Very many words are semantically related (they have the same root in noun and verb or other parts of speech) which results in a high number of words with the same root but different morphological forms.

The difficulty of changing approaches mid-way, between wordlists and semantic domain elicitation

- The strong influence of Swahili (LWC) in some areas
- Participants may give word-for-word translations of compound words in the LWC, only coming back later with a traditional mother-tongue word

- Participants may be thinking, since the LWC is used to conduct the workshop, that it is the language of the workshop, so they see their task as translating, rather than thinking and responding in mother-tongue. Krüger recommends more elicitation by semantic domains, especially early in the process.

- Many of the participants speak Swahili very well and may find it difficult to remember the traditional words of their mother-tongue, once the more common words are covered. People may become overly focused upon translating the exact meaning of the word given in the list and waste time discussing semantic issues—a worthy activity, but not for this workshop.

If people begin with lists, they will feel they have been exhaustive and that their work is done, while there are still thousands of words which haven't been elicited.

Recommendation: Mark passages in the wordlists where additional words can easily be obtained (types of trees, animals, illnesses, or certain activities like working fields, building a house, or cooking special traditional food). Begin with one of these sections in the wordlist. Ask participants to come up with additional words, especially at these marked points. Participants will then be accustomed to the idea of freely giving any words they can think of, using semantic domains.

Preparation for Phonological Analysis

Once collection of vocabulary has taken place:

Step 1. Weed out all compound nouns as well as loanwords and phrases.

Their phonological and morphological structure may exhibit irregular patterns, so they should be studied later, after the inventory of sounds is established. Enter them all in the electronic wordlist, but don't use them for card-sorting. If the linguist chooses to provide parts of speech for the words in the "other" category, this is fine, but for purposes of card sorting for phonological analysis, they will all be placed in the "other" category.

It can be difficult at this early stage to recognize loanwords easily. Especially in Tanzania, many words will be found that are identical with Swahili words. That does not necessarily mean that they are loanwords from Swahili; they might simply be Bantu roots. Sometimes it is possible to ask the participants whether the item in question is a mother-tongue word or borrowed from Swahili. Another hint is when certain sounds (e.g., [l], [f], [D], [T]) only occur in words known from Swahili. In all probability these will be loanwords. When in doubt it is better to **leave possible loanwords in** as they will quickly become obvious throughout the phonological analysis at the workshop.

If the loanwords have already undergone phonetic assimilation to the system of the mother-tongue, they should be left in the wordlist. Also, if there is no mother-tongue word in the language to express the same concept, this indicates that the loanword will need to be expressed in mother-tongue writing, whether it is borrowed or not. Save these words for future reference but they will not be printed on cards.

In addition to weeding out compound nouns and loanwords, one should, at the initial stage, also set aside derived forms, particularly longer verbs with derivational extensions and also the agentive nouns, often formed with the suffix -i, which may cause morphophonological changes.

This means that this early phonological analysis needs to be carried out on a subset of the vocabulary collected. All other words will need to be checked later, to see if their phonological structure fits into the same inventory discovered.

Step 2. Print out the word cards.

Be sure, if several languages will be studied at the same time and place, to use different colored cards for each language's wordlist! As indicated in the software, NOUNS and VERBS will be printed out on cards. LOAN and OTHER (parts of speech other than nouns and verbs, as well as compounds) words will be printed out on paper only, for reference (checking for any additional combinations of letters which may not occur in nouns or verbs). If these occur, they will probably indicate labialization or palatalization of a simple consonant already occurring elsewhere.

Cards will always be segregated by these categories. Use labeled plastic boxes or zipped heavy duty plastic bags to keep them separate throughout Orthography I. In phonological research, one first needs to focus on the Noun Roots and the Verb Roots, i.e., monomorphemic units. Remind participants that we will never want to mix cards from these three categories (unless a gloss is changed and a word is re-categorized). You may also want to give participants a quick review of parts of speech.

Step 3. Generate a phonology report.

Using the Bantu Phonology Tool (the Phonology Template Editor and Search Tool), generate a phonology report (PAW) to produce an inventory of all the sounds and letters of the language as it is currently written.

Print it out, as a reference to be used when checking the inventory of phonemes of the language (to ensure you haven't missed anything) and when filling charts.

Bring these supplies to the Orthography I workshop:

- 2 sturdy containers per language group, labeled "Noun" and "Verb"
- 2 boxes of jumbo paperclips per group
- 1 box of big rubber bands per group
- large tablets of chart paper for all to use
- 1 or more boxes of felt pens for all
- 1 large chalkboard plus a small one for each group, with chalk and erasers
- notepaper
- post-it notes, small

Workshop use of cards and laptops: an overview

While electronic record keeping is efficient and long lasting, laptops should not be visible at the early stages of the workshops. Rather, the "cards" and the target-language speakers should be front-and-center at this point, allowing the target-language speakers' expertise to be the focus. This also allows room for much more insight for speakers of the target language, as a result of the sorting process and a more global beginning to a very analytical process. This means that all changes, either to glosses or the spellings of words, must be immediately written on the cards. As soon as these changes are recorded electronically, a

tick on the card with a colored marker can indicate this has taken place, and a #1 can be typed in the "card" field for a given word in the database. Later on, if a second revision is made, the 1 can be replaced with a 2.

Pre-sorting of cards before daily activities will be a must for the first week or so, due to morphophonological complexities affecting prefix vs. root, and vowel vs. semivowel identification.

The cards are a visual and tactile learning tool. They can be manipulated. Various charts as seen in the worksheet section of the manual can be copied onto large chart paper and placed on a table. Leave card-sized spaces on the chart, where cards can be placed by participants.

Wrapping up, a procedural note: if you can get all your noun and verb entries corrected in the Toolbox database about two days before everyone leaves, you can *then* check the list on your laptop, with the participants watching/assisting. Side benefit: the experts are there with the linguist to not only correct spellings or clarify pronunciation, but to revise erroneous glosses. They learn a lot from the experience, too.

Followup: It may be worthwhile for the TAs to take the time before the workshop begins, to have their cards sorted so that they have at least all the -CVCV roots sorted out, so that the participants don't have to deal with so many variables that first day. And if they have the cards printed up long enough in advance, the other cards can be sorted at leisure over a period of days.

Chapter 3: Orthography Workshop 1: Phonological Analysis

At this point, you are ready to begin phonological analysis. The analysis consists of four parts:

1. consistency checks on what people have written

2. interpretation of the sounds

3. discovery of cases of complementary distribution (through phonotactic research)

4. final inventory of contrastive sounds

Vowels, Week 1

The first week will focus on the vowel system.

Prep (for the week): Read Part 1, chapter 2, "Linguistic Features," Vowel section. Presort for all the C-initial (CVCV) noun roots (V-initial roots will come later). In a Bantu language, finding the roots involves looking beyond the initial noun class prefix. This prefix is usually (V)CV-, but sometimes just (V)- or (V)N-. See Part 2, Appendix J for helpful examples.

Most people who have had schooling in Swahili (and perhaps also via other languages) will initially write their own languages with the five well-known vowel symbols. As stated in the Linguistic Features chapter of Part 1, the language may have five, seven, or nine vowels, or, very rarely, eight!

The general procedure is as follows: Once all -CVCV noun roots and -CVC- verb roots have been identified,[1] the first step is to sort the words by the first vowel, and then check for "same" and "different" sounds, listening to only one vowel symbol at a time. See if the vowels as written on the cards have been written consistently. Is everything the target-language speakers have written as <**i**> the same sound; or is it a mixture of two sounds? (e.g., [i] and [ɪ]). Is all they have written as <**e**> really one and the same sound? Or does their written <**e**> represent two sounds: e.g., [e] and [ɛ], or [e] and [ɪ]?

This consistency checking procedure should be done separately for nouns and verbs. First of all, noun roots have generally two vowels, whereas verb roots have only one vowel. Second, doing it this way would immediately show any systematic differences in vowel distribution between nouns and verbs. In fact, a language was once found in which the nominal system had an inventory of seven vowels, whereas the verbs displayed an inventory of nine!

In a language where prefix vowels are not influenced by harmony with the roots, it is good to begin analysis with an inventory of noun class prefixes, since the participants must be able to recognize them in order to find the roots. However, in some languages, the quality of the vowel in some of the prefixes is determined by the quality of the vowel in the

---

[1]The sorting out of the CVCV nouns and CVC- verbs would really be better done ahead of time by the linguists, I think. The participants won't be ready for deciding which nouns are too long and especially deciding about which are causative or passive verbs (Karen Van Otterloo, p.c.).

root. The quality of vowel used in the transcription of the prefix may actually be a good clue as to the quality of vowel found in the root.

Also, in languages in which Dahl's Law is active, the consonant of the prefix may also change from voiceless to voiced when the root begins with a voiceless C. Thus, the prefixes may have more than one form, even before C-initial roots. TAs should thoroughly familiarize themselves beforehand not only with typical forms of each noun class marker, but also with the phenomena represented within the language group which they will be assisting.

The proposed sequence of events is as follows:

Week 1, Vowels

Lexical ROOTS of nouns are the essential elements for the discovery of the sound system: contrastive vowels and consonants and their allophones, as well as the tonal melodies inherently belonging to the lexical roots.

Vowels Step 1. Discover lexical roots of nouns with C-initial roots

Prep: Pre-sort nouns, looking only for those with -CVCV roots. TAs should reread the Vowel Systems section, chapter 2. They should also look at Noun Class Prefixes in Appendix J.

Comparing the singular form with the plural is very helpful for root identification. Set aside words with a prefix ending in y or w, because the presence of a semi-vowel in the prefix is an indicator of a V-initial root. Example: in the cl. 7 Kifuliiru /ikitala/ 'drying platform' there is a C-initial root: /iki<u>tala</u>/. Contrastively, /ikyusi/ 'smoke', has a V-initial root: /iki-<u>usi</u>/. This causes the prefix to be realized as /iky-/ instead of /iki-/.

Group activity: Since Bantu languages normally have a noun-class system and the citation forms of nouns and verbs consist of a root with its prefix, it is important to first establish the basic forms of the noun-class prefixes as they occur preceding the noun and verb roots (see Appendix J for examples). At this stage, try to identify noun-class prefixes as they occur before a C-initial root, collecting all cards for nouns with the following root structure:

-CVCV   (and later, -VCV and -CV)

Followup: At a later stage, their morphophonological variants will be identified, as they occur before V-initial roots (see Consonants Step 7). It may be helpful to the TA to privately[2] pencil in the tentative forms for each of the noun class prefixes on the first of the morphology charts, **Noun Class Prefixes with consonant-initial roots**, in Appendix J. In order to identify the noun-class prefixes in any typical Bantu language, use the widely accepted numbering system for reference when filling them in on the Workbook Form, Appendix L. See Appendix J for helpful examples and give various alternants as possibilities, so people may recognize the prefixes in their own languages. The importance of this initial morphological research is to establish the roots, which should be underlined at this stage, on the cards.

---

[2]Avoid distracting participants from phonological analysis, at this point. Their noun classes prefixes could have many variants, and these should not be viewed as final yet. Much depends upon the particular vowel system they have, and whether the prefixes contain long vowels, and so forth. At the end of the second week, you'll have enough information to complete the prefix charting, together with the participants.

Vowels Step 2. Chart distribution of each vowel within nouns, in which V1 = V2

Prep: Pre-sort nouns with -CVCV roots in which V1 = V2 . This means sorting the noun data into consistent combinations of vowels: One pile for -CiCi, another for -CeCe, -CaCa, -CoCo, -CuCu, and so on.

Group activity: A speaker of the language should read aloud through each pile, and all should listen to determine whether or not the vowels that are written the same all have the same quality.

If a pile contains two different vowel qualities (though both written using the same symbol), they should be sorted separately, and a second reading should show consistency in each set. We begin with <a> because this tends to be the most "neutral" vowel in a Bantu language, the least likely to pose any problems. In this way, participants can get used to the process without listening hard for allophonic variation or phonemic contrast.

Thus, we take a pile which contains words like eghesara, eghe<u>tanda</u>, en<u>danya</u>, and omo-<u>hara</u> and have someone in the group give the gloss and read the word while the others listen. For each reading, the group must decide:

- whether both vowels in that root sound the same (e.g., does the first <**a**> in <**eghesara**> sound the same as the second <**a**>?
- whether the vowels in this word sound the same as the last word they read, i.e., are the <**a**> sounds heard in <**endanya**> the same as the sounds they heard in <**eghesara**>, or are they different?

Once the pile of <**CaCa**> nouns has been sorted in this way, go on to the <**CiCi**> nouns and the <**CuCu**> nouns, because the high vowels are easier to hear than the mid-vowels. Then move on to the others. Finally, preserve all of your piles of nouns, with rubber bands and post-it note labels.

Followup: Keep these stacks clipped and labeled for future reference. After sorting out all the CaCa, CiCi, CoCo patterns, the OWL knows (or at least has a good first guess at) how many distinct vowels are contained within the data. If the stacks sort into seven piles, there are probably seven vowels in the language, for example. If there are more than five, a temporary symbol should be chosen for the extra vowel, for use on the vowel chart for the next step.

Vowels Step 3. Study vowel combinations, -CVCV noun roots in which V1 ≠ V2.

Prep: Make a chart as shown below, taking into account the vowels of the language. Simply make columns for each vowel of the language across the length of the paper, and rows for each vowel in the second position (with blanks for the identical V slots). Make the boxes big enough to hold the stacks of cards. A partial example, below, for a hypothetical 7-vowel language: Pre-sort nouns with -CVCV roots in which V1≠V2.

Presort the nouns. Take the nouns with di-syllabic roots and sort them according to vowel combinations, such as first syllable: /**a**/ and second syllable /**e**/, first syllable /**a**/ and second syllable /**i**/, for each vowel in the language.

This will show the basic vowel distribution patterns in lexical roots for the language (distribution in the roots cannot be observed in verbs, since basic verb roots only have one vowel). One wants to know the possible combinations of the 5, 7, or 9 vowels, and the combinatory restrictions.

Group activity: Lay the chart paper across a table. Participants read presorted words aloud and place them in their places.

Table 1

V2	V1	a	ɛ	e	i	ɔ	o	u
a			CaCɛ	CeCa	CiCa	CɔCa	CoCa	CuCa
ɛ		CaCɛ		CeCɛ	CiCɛ	CɔCɛ	CoCɛ	CuCɛ
e		CaCe	CɛCe		CiCe	CɔCe	CoCe	CuCe
i		CaCi	CɛCi	CeCi		CɔCi	CoCi	CuCi
ɔ		CaCɔ	CɛCɔ	CeCɔ	CiCɔ		CoCɔ	CuCɔ
o		CaCo	CɛCo	CeCo	CiCo	CɔCo		CuCo
u		CaCu	CɛCu	CeCu	CiCu	CɔCu	CoCu	

Once again, each set of data should be checked for "same" and "different" first, before deciding on the vowel qualities of each.

In a 5V system, sorting like this leads to a total of 25 combinations (5x5), of which 5 were dealt with in the first part, those in which V1 = V2. So a maximum of 20 combinations are left to check.

In a system of 7 vowels, this would lead to 49 combinations (7x7), of which 7 were dealt with in the first session, so a maximum of 42 combinations left. Most likely, not all 42 combinations are found. This study will show if a system of ATR vowel harmony is present in the language (static = root-internal). See Part 1, chapter 2, Vowel and Vowel Harmony sections. Look for combinatory restrictions on the vowel combinations in the roots. If found, the study of ATR vowel harmony will later have to be extended to the dynamic or morphophonological part, namely the behaviour of affixes in combination with roots.

A system with 9 vowels would lead to 81 possible combinations (9x9). However, it is almost certain that such a system displays ATR vowel harmony, and that one can divide the 9 vowels into two subsets of 5 (taking the vowel /a/, inherently [-ATR] also into the [+ATR] set in order to check its occurrence in roots in combination with [+ATR] vowels). This would add up to a total of 25 and 24 possible vowel combinations in roots, of which 9 were already dealt with in the first session, so a maximum of 40 more combinations of vowels in roots are possible.

This study will probably confirm the initial hypothesis of the vowel inventory, if the number of vowel combinations stays within the original set. If not, one needs to do further investigation.

If the number of vowels is confirmed, the group should make immediate grapheme choices, so everyone can get used to the orthographic vowel symbols and start correcting the writing on the cards to reflect their vowel system.

The following has been proven most useful in languages in eastern Africa, and responds well to native-speaker intuition (for further discussion, see Part 1, chapter 2, Linguistic Features, Vowel Systems section).

5V orthography: < i e a o u >

7V orthography: either < i e ɛ a ɔ o u > or < i ɨ e a o ʉ u >

9V orthography: < i ɨ e ɛ a ɔ o ʉ u >

Followup: Go to Appendix L, Workbook Form. Fill in Table 2: **Vowel Combinations** chart, as well as Table 3: **Long Vowel occurrence in disyllabic word roots**, Table 4: **Vowel Length,** and Table 5: **Vowel Harmony.**

Vowels Step 4. Discover lexical roots of verbs. Study vowel distribution in verbs

Group sorting: Sort verbs with -CVC- a roots/stems according their root vowel.

The infinitive of a verb with a -CVC root will have a form like /oko-<u>beb</u>-a/. Sort by the first vowel of the root, as the final vowel in infinitive forms is inflectional rather than part of the root, and is generally /-a/. Each set of items with a particular vowel should be checked for consistency, by people reading and listening. Again, if a pile is divided into two, a second reading should show consistency in each set.

1. Discover the lexical forms of the verbal roots, and underline them.

2. Look for /(V)-ku-/ prefixes, followed by -CVC-a. These are the typical canonical verb pattern. Later, study -VC-a and any other forms.

3. If you find instead /(V)-kw-/ followed by a vowel, put those in a different pile for later, as they contain V-initial roots.

4. Go through each pile of these verbs and decide which pile of nouns the first vowel of the root in each one sounds like. Try to match a pile of verbs to each pile of nouns, but *without* mixing them together.

Followup: Table 1 Vowel Inventory should be filled in at this point, with target-language speakers giving glossed examples for each, though final symbols may not yet have been chosen.

Step 5. Study vowel length

Prep: TAs should review Part 1, chapter 2, Vowels, Vowel Length section.

Group activity: Go through the cards again with the group, listening for vowel length on the roots, both contrastive and compensatory or conditioned.

You will likely find both vowels which have underlying (non-conditioned) length and those which are long due to their environment (i.e., they follow a semivowel or precede a prenasalized stop). Vowels with conditioned length cannot be pronounced short by a native speaker. The lengthening is automatic. Thus, the length of the vowel is not contrastive (cannot make a difference in meaning).

You will probably find words with vowel length which cannot be explained by such environmental conditioning. If this is the case, you will also find words with short vowels, in a phonetically similar environment. This will confirm the presence of contrastive vowel length in the language. Looking through the /a/s for instance, when you find a word with a contrastively long vowel, pull out this card and mark the spelling change on the card (e.g., *mugaata* instead of *mugata*). Make a separate pile (one pile for each vowel) of the words which have a long vowel. Move on to another vowel after completing the first and separate the long /i/s, /e/s, and other vowels in similar fashion. Native speaker perception of underlying length will sharpen as the process goes on.

Note, on a piece of paper (and on a small chalkboard) any minimal pairs (or near minimal pairs) for length. Target-language speakers should probably do the writing on the chalkboard.

Followup: Fill in **Table 3: Long Vowel occurrence in disyllabic word roots**. Also fill in **Table 4: Vowel Length**. Include a gloss for each word of each pair.

Save all of your stacks of cards, sorted by syllable pattern and part of speech, for use next week.

Consonants, Week 2

Prep: Before you begin this week, read Part 1, chapter 2, Linguistic Features, Consonants section. This will help you know what challenges to expect, and it gives recommendations for grapheme choices. The mini-lecture for TAs only, Part 2, Appendix F, Lesson 5: Lesson for TAs, with consonant spelling issues, should be given before this week begins.

The aim of the week is to establish the inventory of contrastive consonants and to study their distribution within -CVCV roots. There are a larger number of consonants, and much of the week will be spent on alphabet matters.

The same principles and approach previously used with vowels apply now to consonants:

- Nouns and Verbs will be looked at separately
- Starting point are initial consonants of the consonant-initial roots.
- The data should be sorted accordingly, and one first listens for "same" or "different," then determines the actual phonetic quality of the consonants.

The approach is like that used for the vowels, except that with consonants it is not necessary to look for words in which C1=C2. We simply look first at each consonant symbol that has been used in the C1 position of the root and sort words by the C1 symbol. Sort on a large chart (see details below) as before, using only nouns the first time and later doing the same with verbs. Since we have kept the noun and verb cards completely separate, this will be simple.

Consonants Step 1. Look at the CVCV noun roots, sorting by C1 (and V1)

Prep: Fold a large, simple chart-paper grid, for use with each language group. Each box on the grid should be big enough to hold a stack of cards and still have the label show in the corner.

Print out the PAW[3] (Phonological Analysis Worksheet or "Bantu Phonology Report"), which lists all segments and where they occur, according to the intuitive transcription documented in word-collection.

Fill in the boxes on the chart paper with every letter or combination of letters identified by the PAW. This should speed the learning process for the day, and ensure that no consonants, digraphs, or consonant clusters are overlooked. For example, without this pre-marking of all consonants, no one will notice that the <mb> and <ng> boxes are empty during the C1 sort (so I changed that part). But they will notice that the NC boxes need to be added when sorting by C2, later.

Large group activity: Part 2, Appendix C, Lesson 2: Helping Readers.

Comparing the PAW to the boxes at the end of this exercise will ensure that no consonant has been overlooked accidentally from the phonemic inventory.

Note before sorting: morpheme breaks and syllable breaks many times don't coincide in Bantu languages! A nasal can be a morpheme in its own right, for example, but when it occurs before a consonant, the nasal and consonant join to form a cluster, i.e., *mb*, *nd,* or *ng*. In nouns, the Cl.9 prefix usually shows up as the nasal in such a cluster. Thus, in a word like *embeho,* the *e-m-* represents the prefix, while the stem is *-beho,* and NOT *-mbeho.* The root-initial consonant is *b*, and not *mb*. **Participants should be briefed on this before the sorting begins**.

Because NC combinations do not appear root initially, during this sort on the initial-C, it is most likely that NC boxes will not appear on the chart (or if boxes are prelabeled with all the segments found in PAW, will be empty). Leave room on the chart to add them in when sorting is done by the second C. This should make it clear to all that NC segments are generally not found in root-initial position!

Group Sorting: Go through each pile of disyllabic nouns (already sorted and paper-clipped) by the first vowel of the root. Disregard the prefix, of course. Noun=prefix+CVCV root. As your group begins to fill the boxes on the chart with cards sorted by initial consonant, write each consonant in the box as it is found (unless boxes have been prelabeled as in step 3 above). The resulting stacks will give speakers a literal picture of the frequency of occurrence of each phoneme in their language. The grid may look something like this (with labels in one of the upper corners, so that they still show when the cards have been placed in the boxes), with stacks of words on top of them, sorted according to the vowel which follows the C1:

Table 2

b	bh	c	mb	t
d	dw	dy	r	ry
rw	r	f	…	

The cards being sorted for initial C are already sorted by the first vowel, so when you finish sorting the first pile, the pile in which V1 is /a/, for example, all the cards on your

---

[3]The PAW is part of the Bantu Phonology Tool, a piece of software which analyzes the phonology of a language and produces a report. It can be found on the SIL website: http://www.sil.org/computing/catalog/show_software.asp?id=116.

consonant grid will be Ca- initial words. When you finish sorting each different pile for C1, paperclip each pile, so that you can keep the *ba-* initial roots and *ka-* initial roots, etc., separate from the *bi-* initial roots and *ki-* initial roots, and so on. Leave each paperclipped stack in its square on your chart paper consonant grid. When you have finished going through all the CVCV noun cards, bundle all the cards on each square together with a rubber band and keep them on the chart paper. Keeping them sorted by first vowel as well as by first consonant allows easy evaluation of whether or not there are significant gaps in the phonotactic distribution—e.g., does a certain consonant sound occur only preceding a certain vowel or set of vowels?

Next, go through all the words in each stack on the grid and listen to them together to make sure that all the Cs which are written the same also sound the same.

When you finish going through all the CVCV root cards and listening to them to check for consistency, you will have an inventory of C1 position consonants. The consonants should immediately be filled in, in a chart with the main points of articulation.

Followup: Examples of each C1 followed by each vowel can now be written down in the Workbook Form chart entitled **Consonants: Root-initial**. Record these examples before moving on to the next step, which involves re-sorting the same cards by the second consonant in the root.

Consonants Step 2. Sort nouns by C_2

A major principle in phonological analysis is "compare what is comparable." In keeping with this principle, the consonant inventory can only be determined by listening to words containing the same consonant in the same position. This is why the second step is necessary: establishing the inventory of consonants found in C_2 position.

Group sorting: Use the same chart paper grid you used for sorting root-initial consonants, but this time to help participants sort the words by the word-medial, or C_2 consonants. Since the words are not presorted by V_2, after sorting the words by the C_2 consonant, go through each pile and sort the words according to what vowel follows the C2 consonant.

For instance, in your stack of words in which C2 is *r*, paperclip together words like *ekekere*, *enkore*, and *ekehore* in one pile, words like *ekemori*, *insuri,* and *embori* in another, and *inshara* and *obhosara* in yet another pile. Then read through, making sure that all the Cs which are written with the same symbol sound the same to the ear of the target-language speaker.

Followup: These consonants will be listed on the Workbook Form chart called **Consonants: Root medial.** There is likely to be a greater number of consonants in this root-medial position, because prenasalized consonants are now included (since the nasal part is part of the root and not part of a prefix). Also, keep track of all labialized and palatalized Cs, so you can list the possibilities at the bottom of the **Root-medial** chart.

Consonants Step 3. Look for complementary distribution of consonants

Group activity: The following steps should now be taken:

1. Check for specific gaps in the CV-distribution (i.e., the set of vowels that may follow each consonant) and see if these gaps would lead to statements of complementary distribution of specific consonants. (The distribution of [l] and

[r] should be checked separately, since their complementary distribution is often determined by the preceding vowel.)

2. Check if there is any complementary distribution based on position within the root, C1 or C2, e.g., [d] only ever occurring as C1 when preceded by a nasal prefix, and [r] only ever occurring as C2 in the root. Such a distribution would suggest that the two are in complementary distribution.

3. Record examples of restricted occurrence or complementary distribution of consonants in **Root-medial**, at the bottom of the chart (and someone can record them for the group on a chalkboard). These will prove very useful to literacy personnel, later.

4. Replace the sorted noun cards in their container.

Help participants to see how restricted occurrences demonstrate allophonic variation. For example, "soft voiced" consonants (fricatives) never will occur after a nasal. They only occur between vowels, while their allophonic variants, voiced stops, will always appear when they are preceded by a nasal. These can often be shown to the group by alternations in the root initial consonant of a class 9 noun when its nasal prefix is replaced by a (V)CV- prefix in a diminutive or augmentative form. For example, the diminutive form of <**imbene**> 'goat' is <**ihihene**> 'little goat' in Kifuliiru. This shows alternation between *b* and *h*, conditioned by the nasal.

Consonants Step 4. Sort consonants in the CVC verb roots, px-C$_1$VC-V

Prep: TAs should reread the Consonants section, chapter 2 of Part 1.

Group sorting: Now, take the CVC-*a* verbs, which were sorted earlier according to their V and follow the same consonant discovery procedure as for nouns, using the same consonant grid chart paper. Sort first for C1, paperclipping the stacks on each box of the grid together to keep the Ca-initial verbs separate from the Ci-initial verbs, and so on.

When sorting for C1 is finished, do a readthrough. People will be listening to be sure that every C they have written has the same sound as the others with that spelling. The readthrough is done to confirm that the sounds represented by a certain symbol are actually all the same phoneme. Just paperclip them by vowel and rubber band each consonant group on the grid. Label as needed with mini post-its.

Followup (with group)**:** If any additional Cs are found after the C1 sorting of verbs (unlikely), add them to the chart of **Root-initial Consonants**, *before* moving on to medial-position consonants and make a tentative decision as to how they will be written, so any spelling changes can be noted on the cards.

Consonants Step 5. Sort consonants in the CVC verb roots, px-C$_2$VC-V

Large group activity: Give Part 2, Appendix D, Lesson 3: Consonant Symbol Choices (if not given previously)

Group sorting: Examine C2 verb roots. You may use the same chart of consonant symbols as the day before.

When the group has finished sorting all the verbs by the consonant in C2 position, read through each stack listening to the consonants sorted by the consonant in second position, to verify that all of those written the same actually sound the same to the target-language speaker. Again, if there are any further consonants discovered, add them to your **Root Medial Consonants chart.**[4] Also note any further co-occurrence restrictions there.

Check your PAW once again just to ensure that no letters have been overlooked. List all the letters, including labialized and palatalized consonants, on scratch paper or chalk board.

It might be helpful to circle the labialized and palatalized consonants while listing them, because they will not be entered on your list of consonant phonemes. You will list them on the Labialized Consonants and Palatalized Consonants sections of the workbook form. These are consonant clusters, and as such will not be represented in an alphabet chart, but the skill of reading them will be taught later in literacy materials. The **w** and **y**, "consonants" in their own right, will probably be represented in any Bantu alphabet, because they will occur in isolation, especially in verbal affixation.

Consonants Step 6. Finalize list of consonant phonemes

Prep: Study "Alphabet Chart Considerations" below.

Study the list of words from the "Other" category. Examine shorter and longer nouns and verbs and words from other categories (pronouns, adjectives, adverbs, numerals, and demonstratives, for example) to see if the vowels and consonants in these words can all be written with the inventory of alphabet letters decided upon. Full analysis of morphophonological issues will be delayed until the morphology workshop.

List any new consonants or clusters which you may (rarely) have found.

Group activity: Fill in the inventory of contrastive consonants, **List of examples for each consonant phoneme**. Target-language speakers will enjoy choosing examples and giving glosses.

Followup: Fill in the charts for **syllable structures**, both word-initial and word-final, in the workbook form. Target-language speakers do not need to discuss these. Word cards are a quick reference.

Consonants Step 7. Study and document V-initial roots of nouns and verbs and vowel harmony (may be done by TA and a few speakers of the language while others work on Step 9)

Prep: In distinguishing true vowels from consonants, study the following examples:

> In the cl. 7 Kifuliiru **ikitala** *drying platform* there is a C-initial root: **iki-tala**. In **ikyusi** *smoke,* there is instead a V-initial root: **iki-usi**, which causes the prefix to be realized as **iky-** instead of **iki-**. It may take a bit of practice with the language to tell for sure what are the C-initial roots, since some roots may not be immediately distinguishable: e.g., in Ngoreme, there are words such as **eghitaro** *well (of water)*, which

---

[4]There is no need to sort by the second vowel here because these are CVC verb roots and the following V will always be *-a* (as long as the TA's have sorted out the passives and causatives and such).

apparently has a prefix **eghi**-, presumable cl. 7. However, there are also words such as **eghogho** *bellows* which could be a cl. 9 or cl. 5 noun with the augment **e**- as its only prefix (**e-ghogo**), or it could be a class 7 noun with an o-initial root (**egh(i)-ogho**). To determine if these two nouns are of the same class or a different class, singular form, while in other languages (probably including those in the Mara cluster) may have a different prefix. Other classes should always have a different prefix in the plural. In the case of these two words, we see that **eghitaro** has a plural of **ebhitaro**, showing that it is indeed a vcl. 7/8 noun, whereas the plural of **eghogho** is **chigogo**, showing that it is a cl. 9/10 noun. So once you know the class in each case, you can see that both of these examples have C-initial roots (Karen van Otterloo, pc). If such a questionable noun has no singular/plural, ask the speaker to add a modifier such as an associative phrase (**ya wengine** *of others* or something similar, and the class will become apparent).

In a language where prefix vowels are not influenced by harmony with the roots, begin with an inventory of noun class prefixes, since the participants must be able to recognize them by now, having underlined roots. In at least some of the Mara languages of Tanzania, the vowel quality of some prefixes is determined by the quality of the vowel in the root. Thus, the prefixes may have more than one form, even before C-initial roots.

Take time in the seven-vowel languages to look at the quality of the vowels in the prefixes and to see how these correlate with the quality of the vowel in the root. This will allow participants to become more aware of what vowel harmony is doing in their language.

Group sorting: Sort all V-initial noun and verb cards separately. List VV sequences as you go, e.g., ai, eo). A speaker of the language can write the list, using a chalkboard or paper.

Look for the shorter roots first: the –VCV roots. Go through the words looking for CyVCV and CwVCV. Further sort these by the quality of the initial vowel.

Followup: Record your preliminary observations on **Table 5: Vowel Harmony** chart.

Consonants Step 8. Identify all noun class prefixes (and variants)

Prep: Gather stacks of nouns, sorted by noun class. Review "Phonological conditioning at morpheme boundaries" in Part 1, chapter 2, Morphophonological Processes.

Large group activity: Give participants a basic presentation on noun classes, tying it in with Swahili and whatever they have learned about noun classes in school. (Optional: make a printout for each participant, showing the standard numbering system along with the sample prefixes of Proto-Bantu and some existing Bantu languages. See Appendix J, Noun-Class Prefixes for Proto-Bantu (Gardner 2005:54).

Group sorting: If this was not done earlier (during root identification), Participants sort through the -CVCV root nouns in their groups, finding words which correspond to the classes they know in Swahili. Go through their cards with them, helping them match up any prefixes not found in Swahili with one of the classes in the numbered list. As they figure these out, consult the lists of both standard and nonstandard noun class pairings

(telling which singular goes with which plural). Record the prefixes for each noun class on Morphology Charts: **Noun class prefixes with Consonant-Initial Roots**) and the form in which they are found when they precede any vowel initial roots, if you've learned this.

Consonants Step 9. Make alphabet charts

Prep: Bring a laptop which has Publisher capabilities and be prepared to use the CD Art of Reading.

Group Activity: This is the time to choose letters to represent each phoneme of the language. Hopefully, most choices can reflect the original, intuitive choices of the people. Regard your alphabet decisions as tentative. It is not unusual to revise symbols later, after more investigation.

Ukwandɨka no Okosoma Ikikiizɨ

Alfabeti ya Kiikizu

The Alphabet

A bit of time should be devoted to alphabetic sequence for each language. There are a number of alphabetic issues which are not automatically obvious to the speakers of the language. These concern the order of symbols not used in a language of wider communication, and also the fact that we are generally listing graphemes for each phoneme in the alphabet listing, including digraphs and trigraphs, and possibly including prenasalised consonants.

The literacy specialist or workshop leader could gradually work out a general alphabet listing containing the sum total of all graphemes used for the alphabets of the different languages. From that "total alphabet," each language group might then deduce its own set. People enjoy this exercise.

Sequence principles: special symbols usually follow the symbol they are most closely related to, such as e…ɛ, i…ɨ, o…ɔ, u…ʉ.

Digraphs and trigraphs, for simple or for complex sounds, should follow the alphabetical order, with secondary sorting on the second symbol, such as m - mb – mv – n – nd – ng – ngb – nj – ny – nz.

Allow 2 to 4 hours per chart for their production on computer (see Part 2, Appendix A: Alphabet Chart Template[5]), once keywords and pictures have been chosen by the group.

Alphabet Chart Considerations

An alphabet chart is a pedagogical tool (but it is no substitute for a primer)

1. If a target-language speaker is already fluent in a language of wider communication, though, such a chart may be all he/she needs in order to begin reading in the mother tongue, especially if the languages are related and they have many corresponding sounds/symbols.

2. An alphabet chart should normally present the phonemes of the language with a simple, direct correlation between sound and symbol.

3. It's not that simple because sociolinguistic considerations should be reflected in the chart. For example, if most target-language speakers of a minority language are literate in an LWC, literacy goals will involve transition from L2 to L1 (or *back transition*).

   • People may want to over-differentiate sounds from their own language if they are already distinguished in the LWC, as in /r/ and /l/ in Swahili. This may actually be acceptable, for purposes of transitional reading.

   • Auditory awareness of the sound/grapheme being taught is key. Usually, it is best for the consonant being taught to occupy syllable-initial position in the root, as people are most aware of sounds in this position.

   • On the other hand, some people may want to show a distinction between their language and the LWC by writing their unique sounds *differently*. The Kwaya people chose to write the soft sound of b, which they say is very different from Swahili's hard-sounding voiced bilabial. For this reason they insist on writing their voiced bilabial (allophonic variants are [β] and [b]), which is always a fricative unless it occurs after an <m>, as <bh>. This choice is not recommended but may be necessary for community acceptance of the orthography (see Part 2, Appendix D, Lesson 3: Consonant Symbol Choices, and Consonants in Part 1, chapter 2, Representing Linguistic Features).

   • The minority language may have absorbed a number of loanwords from the LWC, and hence certain letters from that language, for sounds which are not part of that language's phonemic inventory. Ideally, one symbol should be chosen to accommodate both. Take a look at the list of loanwords which was printed out at the end of word collection. Example: The Tharaka language never needs or uses the letter <s>, using <c> for the voiceless sibilant. But when the New Testament was published, people wanted place names to be spelled using the same place names found in a Swahili Bible, using the letter "s," so two letters for the sound need to be presented in the alphabet chart.

   • Labialization (indicated by <w>) and palatalization (indicated by <y>) can affect virtually any Bantu consonant. The letters w and y also occur alone, allowing them to be taught as separate letters. Again, don't teach mpw, gw, and all the other labialized consonants as digraphs in the alphabet. They function

---

[5]This is a Publisher 2003 file.

more like consonant clusters. They can be presented very simply in a primer, grouped with others, once people can recognize their constituent letters. Once people can decode <k> and <w> individually, for example, a key word such as **kwanda** *'to plant'* can be introduced. This would be followed by an exercise such as the following, where the primary member of the cluster is the consonant, not the semivowel.

- Optimal key words are ones which are not only quickly recognizable, but words in which the learner can easily hear the sound being taught. For this reason, it is ideal for the consonant taught to occur in root-initial position and for a vowel to occur repeatedly, especially in root-initial position.

An alphabet chart is a publicity tool
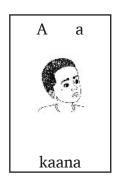
1. It can be a confidence builder, showing the would-be reader that reading the mother tongue is easy.

A   a

kaana

- For this reason, it may be best to keep the chart simple and not teach double vowels as additional letters. In a seven-vowel language, telling people they have 14 vowels can be intimidating and necessitate visual crowding on the chart. If you choose to show the 5 or 7 vowels of the language on the chart rather than 10 or 14 and vowel length is contrastive, be sure to include some picturable words on your chart which *show* this feature, as in the example here.

E   e

mpempe

- The same is true for introducing any prenasalized stops as separate "letters" on a chart (common in Bantu J and E languages). It can be done, though the softer [d] and its harder form in [nd] are actually *allophones*. Since the prenasalized stops (i.e., mp, nd, nc) of Bantu languages normally cannot be pronounced in isolation, they must be taught as digraphs in a primer. Including them in the alphabet chart as well is an option, but not a necessity. If they are *not* taught directly in the chart, be sure to include some key words which have these complex sounds in them, as in the example here.

- Since it is often hard to find certain letters occurring word-initially, try to find words which exhibit multiple occurrences of the letter being taught. The Tharaka language, used in the example above for the letter A, has very few vowel-initial words, so it was decided to use words which contain *only* the vowel being taught, when introducing each vowel. In Bantu words, one can benefit from multiple repetitions of a letter within a word, as in the example key words *kaana* or *mpempe,* shown. For languages with noun-class prefix plus augment, even this may be impossible, but at least the noun prefixes are pronounced with less stress than the root and with less target-language speaker awareness, so they don't seem to interfere with the sound-symbol connection being presented.

2. The chart introduces people to some features of their language's writing system, but *not all.* Word break rules and features of the language's morphology will require further instruction, for would-be writers of the language.

Participants choose key words—picturable words, usually nouns—for their preliminary alphabet chart. All illustrations must be approved by target-language speakers, to ensure that they unequivocally communicate the intended meaning. The linguist helps people choose words for best auditory discrimination of the sound being taught, as described above. A note about illustrations: strive for immediate recognizability

1. Let target-language speakers choose all illustrations, to ensure that they communicate well.

2. If no suitable illustration can be found from the *Art of Reading* CD to represent a given word or letter, a participant may need to draw a picture and have it scanned for inclusion in the alphabet chart, using the template.

Unresolved alphabet issues

Any unresolved phonological issues such as choice of some alphabetic symbols, ATR harmony, semivowel/vowel problems, or vowel length should be noted in the manual Workbook Form, as some deeper and more detailed analysis will be necessary. They may require focused attention at the next workshop, including orthography testing. Inform the literacy consultant or the linguistic consultant as to these issues. Provide examples.

List Revision and Text Collection: ongoing activities this week and next

Group activities: (These can be done contiguously with card sorting, by subdividing a language group for focus on different tasks)

1. Begin revising the electronic wordlists, based upon decisions made.

2. Record 3 to 4 short stories (less than 5 minutes each) on a sound recorder, creating wav files.

3. Teach some target-language speakers to transcribe the stories, using Speech Analyzer.

   • Each story is written using black ink, one line per sentence or clause.

   • Below each line of the story, in red ink, is a word-for-word transcription in a LWC

   • On a third line, if needed, transcription can be done in blue ink, with more natural sentence flow in the LWC.

4. Transcription reminders:

   • listen for pauses. Each one signals a new line on the paper.

   • mark them with full stops

   • leave room in the margins

- use capital letters at the beginning of sentences (or phrases, at this point)

- leave three lines below each transcription

- choose one color of ink for the whole story

- it's helpful to number the original sentences

- Check any LWC glosses which need revision to match the vernacular word on the card/computer.

- Take care that transcription of the gloss is as close to the vernacular morpheme order as possible. For example, if the verb in the vernacular has an object marker in it, the word-for-word LWC (a Bantu language) translation should have a corresponding object marker. People will need a little lecture here, with examples of "free" and word-for-word translation. Use a language like English, which is guaranteed to be quite different from theirs. "Once upon a time there were three bears…: "Once upon a time" would literally be "Mara moja juu ya wakati," whereas, written freely, it would be "Hapo zamani" or something.

Tone, Week 3

Try to find a place where each language group can work without distraction from the other groups, all of whom will be whistling like a flock of birds. Two or three target-language speakers are an optimal number. If there are more, some may be best used at this time for recording or transcribing stories in another room or choosing illustrations for an alphabet chart.

While complete tonal analysis is impossible at this time, a number of basic questions can be answered now:

- Is tone on noun prefix stable (i.e., always high)?
- If it changes, does it change only when followed by a V-initial stem, or even when followed by a C-initial stem?
- Tone on roots: are there rising and falling contour tones on the surface? In what syllable positions?
- Can a contour be found only on a long vowel, or on short vowels as well?

In most Bantu languages in Central and Eastern Africa, the basic tone system consists of two underlying tones, High and Low. However, there are exceptions here and there.

Tone Week Step 1. Study tone on verbs

Since tone is a relative feature and always perceived in relation to tone in other words, the tone of a particular word is usually elicited within a frame. A frame for verbs could be the imperative followed by an adverb, like "_____ well!" This would work for most intransitive verbs and others which don't require a stated object. A frame may be essential in West Africa, where floating tones cause perturbations on the tones of adjacent words. Since this initial charting activity is just giving the big picture though, a frame is optional

Prep: Use chart paper to track tone patterns on verbs, like the one for nouns below, but without the prefix column.

Group sorting (while others are telling or transcibing stories)**:** Set aside all verbs which don't have C-initial roots.

Someone in the group reads each word and whistles all verb infinitives of a particular syllable structure. Divide the melodies into consistent groupings. It is most likely that verbal infinitives will fall into two tonal classes at most, and perhaps only one. The longer a verb stem is, the easier it is to hear the different melodies, so include not only the CVC-FV verb roots, which were sorted already for segmental phoneme discovery, but also the C-initial verb roots with extensions, (i.e., *okorareka, okoraghora, okobharoka*).

If the root is C-initial, have someone read it and whistle its tonal melody, so that each syllable (of the whole word, not just the root) gets a whistle and make piles of "same" and "different" melody. This will automatically result in the sorting of two-, three- and four-syllable verbs into separate piles, since HHHL is not the same as HHHHL. If there are different tonal classes for verbs, you may have *two* piles of five-syllable infinitives with different melodies.

Do not make separate piles for differences in tempo! Ignore the length of vowels, listening only for differences in pitch pattern among verbs with the same number of syllables.

The first possible expectation would be to have two groups: H-tone verbs and L-tone verbs (whereby the H or L tone belongs inherently to the root, but may be realised on the root itself, or later, or sometimes even earlier). Another possibility is predictable high pitch on the antepenultimate mora.

Listen to the whistles and try to mark down the tonal pattern of each group of verbs. With practice participants often recognize the tonal group to which a verb belongs, with lots of whistling and listening. Pitch traces are the best way to record the patterns. Most Bantu languages have downdrift, so that a H tone that occurs after a L tone will have a lower pitch than a H tone that preceded the L, and a pitch trace transcription shows this more accurately.

If there are tonal classes for verbs, note this on the **Tone** chart of the Workbook Form.

Tone Week Step 2. Study tone on nouns

For a first impression of nominal tone in most Bantu languages, just elicit nouns in isolation. If you are in a language area in which floating tones or H spreading are common, though, use a frame. Try a verb plus noun frame, such as "Look at the _____." This would result in a verb preceding the noun. If no perturbations seem to occur on adjacent words, try a simpler frame such as a noun plus demonstrative, "That ____."

There is a slight complication with preceding frames for languages where there is a vocalic augment. The augment of the noun will coalesce with the final vowel of the preceding word. Also, in languages where the adjective has an augment as well, it may be hard to pronounce without the augment coalescing with the final vowel of the noun. Instead, using a following associative phrase may be helpful. Regardless, when augments are involved, elicit the isolated noun first, and then use the frame.

Prep: Sort all nouns of the basic -CVCV root structure by noun class, and design a chart for nominal tone. Make the boxes big enough to "hold" stacks of word cards, sorted by melody (written with pitch traces at the top of each box) and number of syllables in the root.

Group sorting (while others are recording or transcribing stories): Sort the nouns by class, and then by roots. Stick to -CVCV and -CVCVCV roots. Then,

1. Go through each stack using the same procedure as for verbs, reading and whistling each word. Only sort by melodies in which the pitches show a different pattern, and between different numbers of syllables. The speakers of the language should get into a rhythm of speaking the word (held up or written for all to see), then whistling it, then creating groups of words which have the same tonal melody.

2. The TA writes a pitch trace for each pattern in a box of a chart like the one below, made on large chart paper. One participant should be given responsibility for keeping a particular stack or stacks neatly piled and/or clipped together. Based on the checking system explained above, list the tonal melodies for nouns. If there is enough data, this should be done preferably per noun class (so that one can ensure the full range of possibilities is found in every noun class).

3. As cards are matched with a pitch trace, stacks are formed.

4. Groups listen for the pattern of words within each stack, for consistency of melody.[6]

Most patterns (or all) should show up in all the noun classes. However, by keeping the different noun classes in separate piles, you will have a pile of cl. 1 nouns with a LLHL pattern and a pile of cl. 3 nouns with a LLHL pattern, a pils of cl. 5 nouns with a LLHL pattern. Have a spot ("square") for each pattern, keeping the cards of each class paperclipped together on that spot.

Most Bantu languages will have four or fewer classes of tone patterns in nouns, and the same patterns will be found in all the noun classes. However, there will probably be some exceptional words that do not fit easily into any group. Do not expect to have a thorough analysis of tone at the end of this exercise, but it is hoped you will be able to document some basic tonal melodies.[7]

---

[6] If there is an occasional multi-syllable word that does not fit any patterns, it should be set aside for later evaluation. It might easily prove to be a compound word or a phrase.

[7] In some countries, such as Tanzania, where the influence of another language is very strong, the language may be in the process of losing tonal contrast. Be aware. If you find no contrast between words in isolation, it may be useful to test the words within some frames, i.e., short phrases, to see if there is a contrast which emerges in context.

Tone Chart

Syllables	At the top of each square below, (as needed) write pitch traces for sorting cards into tonal patterns. Additional columns may be added to the right for causative/passive if the melodies are different.			
Prefixes + 1 syllable,C-initial	Ex: [⎯⎯⎯﹨] Put cards here	Ex: [﹍﹍⁄] Put cards here		
Prefixes + 2 syllable,C-initial				
Prefixes + 3 syllable,C-initial				
Prefixes + 4 syllable,C-initial				
Prefixes + 5 syllable,C-initial				
Prefixes + 2syllable V-initial				
Prefixes + 3syllable V-initial				
Prefixes + 4 syllable V-initial				
Prefixes + 5 syllable V-initial				

At the end of the process, keeping the different noun classes in separate piles, the result will be a pile of cl. 1 nouns with a LLHL pattern and a pile of cl. 3 nouns with LLHL pattern, a pile of cl. 5 nouns with the same pattern, and so on. Each class will be paper-clipped, and all of the same melody and number of syllables will then be stacked and rubber-banded together.

Do the same exercise for nouns as previously done for verbs. The most probable finding will be two tone groups: H-tone verbs and L-tone verbs (whereby the H or L tone belongs

inherently to the root, but may be realized on the root itself, or later, or sometimes even earlier). Another possibility will be predictable high pitch on the antepenultimate mora. Based on the checking system explained above, list the tonal melodies for nouns. If there is enough data, this should be done preferably per noun class (so the full range of possibilities is explored for every noun class). Similarly, the tonal classes for verbs can be established.

Tone Week Step 3. Record function of tone, finalize wordlists and alphabet charts

Prep: These activities are primarily to be done by a TA. Keep a complete list of minimal pairs: nouns which are segmentally identical but which have different tonal melodies, and similarly verbs which have different underlying tones.

Since there are generally four different noun-tone patterns, and only two different verb-tone classes, there are generally more lexical minimal pairs in the verb system than in the noun system. The function of tone in the grammar is generally much more important.

It is very important to find out where the tone functions in the grammar. One can try to find out if it functions in particular domains which have tonal contrast in other Bantu languages, typically:

- 2sg. vs. 3sg. subject prefix *u-* with L and H tone underlyingly (but sometimes, because of tone shift, not even realised on the morpheme itself, but a couple of syllables to the right)
- 3sg. vs. 2pl. object prefix *-m-* with L and H tone respectively
- Verb tenses, esp. 1st degree past vs. 3rd degree past
- Relative clauses

Followup: Note the number of tone classes for nouns on the **Tone** section of the Workbook Form, with examples.

List the tonal melodies for nouns. If there is enough data, this should be done preferably per noun class (so that one can ensure the full range of possibilities is found in every noun class). There is room in the tone worksheet to note any observations or impressions about such features and processes as:

- Depressor consonants
- Tone spreading
- Tone shifting
- Downdrift and/or downstep

Tone Week Step 4. Edit and review

Prep: Bring laptops today. This can be a day for bringing closure to either tone classes, alphabet charts, or text recording and transcription. It is also an excellent time for a few target-language speakers to assist the TA in correcting spelling and glosses in the electronic data, referring to the corrected word cards.

Small group activities: If you can get all your noun and verb entries corrected in the Toolbox database about 2 days before everyone leaves, you can then check the list on your laptop, with the participants watching/assisting. Side benefit: the experts are there with

the linguist to not only correct spelling or clarify pronunciation, but to revise erroneous glosses. The participants learn a lot from the experience, too.

Assign someone from each language group to prepare, together with the rest of the group if possible, a 10-minute report on the linguistic features of their language (i.e., how many vowels do they have, anything interesting about their noun class prefixes, their consonants, tone, or length…) and interesting features of their alphabet. The report will be presented on the following day.

Finalize the 500-wordlists and alphabet charts for each group. Print out[8] at least one copy of each, and allow one or two people within each group to check through them, while the others work on tone or story transcription. It is suggested that charts and lists give credit to the people who checked and developed them, with a listing of the names of the workshop participants somewhere on the papers.

Tone Week Step 5. Close workshop

Large group activity: This is the day for:

- review of what has been learned: a representative of each language group should make a short presentation to the large group about the linguistic discoveries relevant to their language.
- closing ceremony.
- distribution of a limited number of alphabet charts (one per participant) and wordlists, which then may be viewed by the larger language community after people leave.

Before alphabet charts are distributed, instruct participants: "Target-language speakers need practice reading their own language and may not at first like all the spelling choices we have made here. This is natural. Some things will not be written quite the way they expected. But if many people have trouble reading the same thing, write that word down on a separate piece of paper, and tell us next time. If people say, 'We don't say it like this. It has a longer sound,' what will you (the former workshop participant) tell them?" Be sure they are prepared to answer such questions.

Distribute, read together, and discuss the following instruction sheet, translated into the language of wider communication. It will accompany the wordlist and alphabet chart.

---

[8]Use Lexique Pro to print the wordlists. Someone will need Publisher 2003 in order to print the alphabet charts.

Things to do each time you show someone the wordlist: informal testing of preliminary orthography (Translate)

Show both the chart and the wordlist to at least 15 people. When you do,

1. Ask them to sign their name on the back of the list.

2. Show the alphabet chart first. If there are any new symbols on it, make sure people understand them.

3. List any problems with alphabet chart pictures (or suggestions), if any.

4. List (later, when they're not looking) any words which are often a problem for people to read.

5. If someone to whom you showed the list speaks another dialect of your mother-tongue, please note this next to their signature.

6. Were there any dialect-related problems anyone had, either with the ABC chart or the wordlist?

7. Record any complaints.

8. Evaluate before you come back:

   • Did most people like the ABC chart?

   • Did most people like the spellings of words in the list?

   • Do most people like the idea that their language is now written?

Chapter 4: Orthography Workshop 2: Morphophonology

Morphophonological processes in Bantu languages involve:

(C)V – V hiatus resolutions: hetero-syllabification (e.g., formation of two syllables from one), elision, and glide formation (often two or three of these solutions are used in one language)
C – V morphophonology
C – C morphophonology (particularly in cl. 9/10 noun-class morphology, and 1sg subject/object prefix morphology)

In Bantu languages, the study of morphophonology should be divided into two parts:

Nominal morphology and morphophonology
Verbal morphology and morphophonology

The topics under nominal morphology and morphophonology are:

Noun-class prefixes of nouns in their citation forms
Class prefixes used with adjectives
Class prefixes used with numerals
Class prefixes used with possessive pronouns
The associative markers
Class prefixes used with demonstratives
Class prefixes used with any other noun modifiers

The topics under verbal morphology and morphophonology are:

Inflectional
- subject prefixes (for participants and for classes)
- object prefixes (for participants and for classes)
- relative prefixes (for participants and for classes)
- T/A markers
- Negative markers
- any other inflectional elements particular to a language

Derivational
- Verb – Verb:
    - productive derivational suffixes: causative, passive, applicative, reciprocal, for example
    - combinations of these derivational suffixes
    - changes in verbal aspectual suffixes because of the use of (combinations of) derivational suffixes
- Verb – Noun: agentive nouns derived from verbs
- Verb – Adjective: adjectives derived from verbs

Once clear and systematic solutions have been found for these various morphophonological issues, charts can be drawn, particularly of the class prefixes functioning with the various nominal modifiers, which often contain a lot of morphophonology. Record not only orthographic conventions chosen, but the actual pronunciations, when they diverge from

spellings! It is important that a record be left behind for any future linguistic research. For example, the Kwaya language of Mara Tanzania has two charts for associative marker: one reflecting the spelling decisions, and the other representing the phonology of the language.

Some charts may later be distributed, or hung up on the walls of a translation office, so that people frequently refer to them and learn to write correctly.

By way of example, see the following chart. It portrays the complete set of possessive pronouns for all classes in the Yaka orthography. The language is YAKA (C10), CAR.

		1sg	2sg	3sg	1pl	2pl	3pl
		-amu	**-ɔfɛ**	**-ɛi**	**-ɛsu**	**-ɛnu**	**-aɓɔ**
1	**moto**	**wamu**	**wɔfɛ**	**wɔi**	**wɔsu**	**wɔnu**	**waɓɔ**
2	**ɓato**	**ɓamu**	**ɓɛfɛ**	**ɓɛi**	**ɓɛsu**	**ɓɛnu**	**ɓaɓɔ**
3	**mole**	**wamu**	**wɔfɛ**	**wɔi**	**wɔsu**	**wɔnu**	**waɓɔ**
4	**mele**	**myamu**	**mɛfɛ**	**mɛi**	**mɛsu**	**mɛnu**	**myaɓɔ**
5	**dɔkɔ**	**dzamu**	**dzɛfɛ**	**dzɛi**	**dzɛsu**	**dzɛnu**	**dzaɓɔ**
6	**matɔkɔ**	**mamu**	**mɛfɛ**	**mɛi**	**mɛsu**	**mɛnu**	**maɓɔ**
7	**eɓuku**	**yamu**	**yɛfɛ**	**yɛi**	**yɛsu**	**yɛnu**	**yaɓɔ**
8	**ɓeɓuku**	**ɓyamu**	**ɓɛfɛ**	**ɓɛi**	**ɓɛsu**	**ɓɛnu**	**ɓyaɓɔ**
9	**mboka**	**yamu**	**yɛfɛ**	**yɛi**	**yɛsu**	**yɛnu**	**yaɓɔ**
13	**lofuma**	**lwamu**	**lɔfɛ**	**lɔi**	**lɔsu**	**lɔnu**	**lwaɓɔ**
14	**ɓokia**	**ɓwamu**	**ɓɔfɛ**	**ɓɔi**	**ɓɔsu**	**ɓɔnu**	**ɓwaɓɔ**
19	**fifuma**	**fyamu**	**fɛfɛ**	**fɛi**	**fɛsu**	**fɛnu**	**fyaɓɔ**

Orthography Sketch and Writers' Guides

With the information obtained so far, a writers' guide can gradually be developed. The writers' guides are not a phonological justification for spelling decisions. They present the agreed-upon spelling, with examples, to provide guidelines for correct writing for present and future generations. Examples of writers' guides are given in Appendix 1, and others may be available from your entity.

The following elements should definitely be found in a Writers' Guide:

- choice of alphabet letters
- word boundary decisions
- capital letters, punctuation rules

Nouns, Week 1

Step 1. Begin charting adjectives

Prep: Literacy Lecture, Part 2, Appendix G, Lesson 6: Goals and Products of Workshop 2 and charts described below and pictured right. Read Part 1, chapter 2: Morphophonological Processes, Principles of Tone Marking, and Word Boundary Principles.

Pedagogical Materials: Large wall chart copies of each language's noun classes, with prefixes and an example noun for each. These charts will be used almost daily as a reference.

Poster-size checklist of charts to be filled and consultant-checked during the 3-week workshop, showing all language groups' progress (see below).

Linguistic Lecture: rules for vowel length, Dahl's Law, noun phrase constituents, adjectives.

Group activity: Chart adjectives which begin with consonants, and those which begin with vowels, for each noun class.

Step 2. Chart noun phrase numerals

Prep: Put big wall charts up, attaching 2 blank chart papers to the right of the noun class charts. Linguist demonstrates how to fill in the first 4 rows of one language group for all, before dividing into groups.

Group activity: Chart numerals, 1–10

Followup: Fill in **Numerals** chart.

Noun class 3	Noun	Number 1
(Ngoreme)	omoti	gamwe
(Kwaya)	omugongo	gumwi
(Simbiti)	omote	gomwe
(Ikiizu)	~~umuti~~	~~gumwi~~

Some noun class prefixes will not be retained beyond the number five. The enumerative for class 10 is always a noun.

Linguist lecture: As a review and a learning experience, a linguistic consultant charts samples from all, on flip chart, as above: Noun Classes 2 and 3 with numbers 1, 2, 6, and 8 in separate chart pages.

Step 3. Chart demonstrative phrases

Prep: Read Word Boundary Principles section of Part 1, chapter 2.

1. Linguist explains how "near" and "far" work in Swahili, with a chart:

(singular)	huyu	yule
(plural)	hawa	wale

2. Linguist demonstrates the same thing for another Bantu language. See **Demonstratives and Demonstrative Phrases** chart, as a formatting guide for near to speaker, near to hearer, far from both, and previous referent (but known about mutually).

Group Activity: Chart demonstratives on a large wall chart, aligned with noun class prefixes.

Followup: Copy the wall chart onto **Demonstratives and Demonstrative Phrases** chart.

Step 4. Chart demonstrative forms: other

Prep:

Group Activity: Chart **Demonstrative Forms: Other** on chart paper, with the group.

Look for Emphatic forms to show up, and note them if they occur.

Followup:

Step 5. Chart interrogative forms

Continue in the same manner with interrogatives and with all of noun phrase charting below.

	Nomino: Ngoreme	
	kiambishi	mfano
1	omu-	omugabo
1a		yaya
2	aba-	abagabo
2a	ba-	bayaya
3	omo-	omokune
4	eme-	emekune
5	ri-	riguha
6	ama-	amaguha
7	egi-	egikoba
8	ebi-	midomo
9	en-	engoha
9a	e-	ehoonyo
10	chin-	chingoha
10a	chi-	chihoonyo
11	oru-	orunyanki
14	obo-	obore
15	ogu-	ogutereri
16	aha-	ahase

Step 6. Chart pronominal forms-1

Step 7. Chart pronominal forms-2

Step 8. Chart pronominal forms-3

Step 9. Chart diminutives and augmentatives

Step 10. Chart possessive pronouns

Followup: Fill in electronic charts for each of the following, after filling in blank chart paper with participants in each language group. Be sure to record any spelling

decisions with examples and any unresolved spelling issues (with examples) on the Spelling Rules page following the chart.

Step 11. Locatives

Prep: TAs should read Part 1, chapter 2, Morphophonological Processes.

Literacy Lecture: "What is a word?" Part 2, Appendix H, Lesson 7: Word Break with locative examples (and their 3 noun classes) from Niha.

Three possible orthography choices are offered: 1) writing the underlying locative, with vowel changes written only on the following noun with a space between the two words; 2) writing the vowel change onto both the locative and the following noun, allowing vowel length to be accurately reflected; and 3) writing the vowel change only on the locative, dropping it from the prefix position on the noun and using a hyphen.

Group activity: People break into groups to chart their locatives, written disjunctively in one of these ways. Experimentation is good before a decision is reached.

Followup: Record both the spelling decisions and the phonological processes on the charts. This can be done by someone assigned the task.

Step 12. Chart associative constructions

Linguistic or Literacy Lecture: A consultant gives a description of the procedure for filling in the chart of associatives. She gives Bantu examples of the morphophonological changes across the word break. Options for writing them do *not* include hyphenation (literacy consultant provides all the possibilities). We don't want associative markers to look like locatives, stuck onto the beginning of the noun. Also, since in Swahili the associative is written disjunctively, people should not object to writing them separately from the nouns they precede. People will not usually question the suggestion. The waIkiizu (E402), however, chose to just write the associatives and the locatives in their underlying forms. After all, they have 7 vowels, and the variant speech forms are many indeed. They also are convinced that the associative is always *wa*, despite its morphing before every V-initial noun.

Followup: Record both the spelling decisions and the phonological processes on the charts. This can be done by a person assigned the task.

Verbs, Week 2

Step 13. Chart Copular Forms

Step 14. Chart Copular Forms Part 2

Step 15. Chart Morphophonological Processes, Perfective

Step 16. Chart Verbal mood: Imperative Plural

Step 17. Chart Verbal mood: Imperative Singular

Step 18. Chart Verbal mood: Subjunctive

Step 19. Chart relative phrases, Object Relative and Subject Relative

Small group: Linguist or TA presents several possible forms of relative phrase. In each case, an element in the verb phrase coordinates with the noun at the head of the phrase. The phrase functions like an adjective modifying the noun, so you still need a "real" verb later in the sentence.

Step 20. Chart Use of the Augment on Nouns, continue charting verb forms

Step 21. Check charts with a linguist

Prep: Post a progress chart like the following up on a wall, minus the *Xes*, for all to see. Groups may cross off charts after they have been "handed in" electronically to the consultant/s. Consultants circle the *Xes* on the wall after charts have been checked with a target-language speaker.

Wrapping up Charts

Chart Needed	Ngoreme	Kwaya	Ikizu	Simbiti
Nouns: Demonstrative forms- other	x	x		x
Demonstrative/Demonstrative Demonst. Phrases	x	x		x
Interrogative	x	x	x	x
Pronominal forms -1	x	x	x	x
Pronominal forms -2	x	x	Help	x
Pronominal forms -3	x	x	x	x
Adjective	x	x		x
Associative Constructions	x	Help		x
Diminutives and Augmentatives	x		x	x
Enumeratives	x		x	x
Locatives	x	x		x
Possessive pronouns		x		x
Verbs: Copular forms				
Copular forms part 2				x
Morphophonological processes…Perfective				
Verbal Mood: Imperative plural	x	x		x
Verbal Mood: Imperative singular	x	x		x
Verbal Mood: Subjunctive		x		
Verbal Mood: Subjunctive 2		x		x

Verbal Prefixes: Subject.		X	X	X
Verbal Prefixes: Subject 2				X
Verbal Prefixes: Object	X	X	X	X
Verbal Prefixes: Object Relat.		X		X
Verbal Prefixes: Subject Relat.				X
Verbal Tense: Future 1		X	X	X
Verbal Tense: Future 2		X		X
Verbal Tense: Future 3		X		X
Verbal Tense: Future 4		---		X
Verbal Tense: Past 1		X	X	X
Verbal Tense: Past 2		X		X
Verbal Tense: Past 3		X		X
Verbal Tense: Present or unmarked				X
Verbal Tense: Present 2	---	---	---	---
Verbal Tense-Aspect/other				see Past 1-3
Verbal Tense-Aspect Anterior	X	X	X	see Past 1-3
Verbal Tense-Aspect Past Anterior			X	see Past 1-3
Verbal Tense-Aspect Progressive	X	X		X
Verbal Tense-Aspect Present Persistive		X		X
Verbal Tense-Consecutive				X
Vowel Harmony in Extensions				X
Conditional (if it exists)				X
Regret?				X
Stories (Hadithi)	X	X	X	X
Adjectives (Vivumishi)	X	X	X	X
Associatives (-a Unganifu) (optional)	X			
Noun Classes (Ngeli za Maneno)	X	X	X	X

Verbs and checking of charts, Week 3

Prep: Review Word Boundary Principles in Part 1, chapter 2, Linguistic Features. This week has three goals for each language group:

1. filling and checking of all remaining morphology charts

2. review of spelling rules (i.e., word break rules, writing of grammatical vowel length, spelling of copulas and focus markers, grammatical or lexical tone marking) decided upon as part of the charting process

3. editing an oral text which was transcribed at the previous workshop, to be published

Charts to be checked, with linguist (she/he has the master set of all charts from this point):

• Associative

• Locative

Items to be published:

• Noun class charts, copied off from worksheet and corrected if needed

• Adjective charts

• Verbal prefixes: subject or object

• Stories

How chart checking should be done:

• Linguist sits at the front table with a laptop. People bring their assigned chart to the linguist on a flash drive. Bring a target-language speaker along. TA is optional.

• Linguist asks questions and types phonemic information and linguistic details as needed on each "chart" given to him/her. Linguist highlights any spelling questions which arise, consulting the Orthography Manual, testing, and/or discussing challenges with the literacy specialist or consultant.

Literacy consultant: Make sure to get alphabet charts checked off, vowel books (for 7-vowel languages) approved. Assist with orthography questions and story editing issues. Make sure you have an edited story from each group by two days before the workshop ends. Collect all four items for publishing.

Group Activity: Follow the rules the group have already chosen and edit their stories. Now try reading them a couple of times. If rules need to be changed or added, call a consultant over. If they do, some charts will have to be rewritten.

Encourage groups to turn their charts in daily this week and to have them checked by your linguistic consultant working with a target-language speaker, to ensure that the charts accurately represent the pronunciation, not just the orthographic conventions chosen, of the linguistic phenomena recorded. The literacy consultant should offer alternatives and testing suggestions for any unresolved orthography issues as they arise.

Check these charts with linguist as time allows:

• Imperative
• Noun class charts, copied off from worksheet, and corrected if needed
• Adjective charts
• Verbal prefixes: subject
• Use of the augment (questionnaire)
• Copular forms
• Imperative (singular?)

- Relative phrases
- Vowel coalescence chart, if not already completed

Step 21. Print charts

- Print all charts, copy all electronic charts available, backing them up onto the workshop leader's computer
- Print all ABC charts which need a reprint
- Publish any booklets which may be needed

Step 22. Edit stories in groups, referring to the writing rules the group chose during group charting

Large Group activity: Story Edit (find in Appendix I, Lesson 8, page 137) Review the writing rules for one language group, beginning with a sample sentence or two from one of their stories. Outline a simple poster chart, which will be filled in by each group, listing rules for spelling:

- Associatives
- Locatives
- Copulas and other clitics in the language (when analyzing, pay particular attention to varying functions or possible similarities or contrasts with a focus marker /n/ and with the conjunction /na/)
- Tone
- Other

Small group activity: Groups now begin listing the rules they have already prescribed, as a review.

Groups begin editing their stories (for large groups the stories have been written out on chart paper, copied exactly as they were transcribed during Orthography 1). The entire group should be able to see the stories and editing should be a group process. Go through the stories at least twice, making sure that all locatives and associatives have been noted.

Ask different people to read the story aloud. If several people have problems at the same place in the story, note where they stumble and inquire as to the reason. Get the story checked by a consultant and prepare it for publishing.

If time allows, participants may select an illustration for the story from the *Art of Reading* CD (see www.ethnologue.com/LL_docs/intl_illustrations.asp).

Step 23. Closing ceremony and distribution of printed materials

Large group activity: During closing ceremony, the chosen story is read to the whole group, and each participant is given a copy of the 3 charts and the story, plus any edited alphabet charts or vowel booklets.

Appendix A Alphabet Chart Template

Title in local language
Alfabeti za Ki…

A a yagakana	B b baaba			

Appendix B

Lesson 1: Goals and Products of Workshop 1

Lecture presented by Leila Schroeder in February 2007

Learning Objectives for this lecture:

1. People know what tangible results to expect from the workshop

2. People understand that choice of grapheme to represent a sound varies with the needs of a language (so sociolinguistic factors will not have a larger influence on grapheme choices than optimal for reading/writing)

The lecture:

We are here to begin developing writing systems for your languages. We hope you will leave here with an alphabet chart and a list of words written in your language. This work will not be easy. Your biggest challenges will be:

1. Helping writers by making spelling easy for them, with simple rules

2. Helping new readers by using one symbol for each sound in your language whenever possible. We use alphabets to do this.

   Show *sample alphabet charts (i.e., for related languages). *Chalkboard needed

   We hope that each of you can leave here after three weeks, with a chart like one of these, which helps people be aware of all of the sounds which make a difference in meaning for your own language. Sometime during the second week of the workshop you will begin to choose pictures of things which everyone in your language group can recognize, to represent the sounds of your language. An ABC chart like these, but with pictures of words from your language, will help people learn to read it someday.

   You will be working very hard for three weeks on things which are *new* to you, yet also *familiar*. They are familiar, because you use these sounds and meanings every day. They are also new and strange, because you have not had to think much about how they would look if they were written down.

   Writing systems teach people many things about their language, things they have never had to think about until they are written down. (Distribute sample alphabet charts.) What can you learn about a language (and a culture) by looking at its alphabet chart? (Discuss briefly.)

   See the Tharaka alphabet chart (for example). How many vowels does this language have? Do they just have short vowels, or long ones too? Look closely. Any other observations? Some of the Tharaka "letters" are very long. Why have we put pictures in the chart, rather than just words?

Look at Suba. This language is more closely related to yours. Are most of the items familiar to you? The Suba alphabet chart teaches the long and short vowel sounds separately, though they actually have the same sounds and distinctions as Tharaka vowels.

Look at the letters.

- Suba has the letter L. Does Tharaka? No, because in Kitharaka the two letters <r> and <l> do not make a difference in the meanings of their words.

- On the other hand, Tharaka has the letter û and the letter î, 7 vowels in all. Does Suba? No, Suba doesn't need them because they only have five vowels which make a difference in the meanings of their words.

- Which alphabet has a z?

- Which one has an s? Why do you think Tharaka chose c instead of s? There are problems with their choice for spellers and people who read more than just Kitharaka, but English or Swahili. We will talk about such things later.

- Which letters do you think are easiest to read? Why?

- Which letters might be the hardest, both to spell and to read?

We will be talking, next time, about the things which make an alphabet easiest to use, both for recognizing different letters and for writing them in words. As we study your languages together, we will all talk more about the things you must consider in choosing letters for your language's alphabet.

3. Helping new readers by breaking up the sounds into words and sentences.
   Example: sinunuimkatemjinikwasababuhaufai
   Exercise: Break into words—Sinunui mkate mjini kwa sababu haufai.

   When we talk to our friends, we usually drop some of the sounds in words, knowing they will understand us. In English, when we talk fast, we say "G'bye." But if we speak slowly, we are actually saying "Good Bye." That is the way we need to write it: "Good Bye." We native English speakers say, "Dijuseeimatth store?" But when we write, we break these sounds of fast speech into slow speech, and really help the reader and writer to understand: "Did you see him at…" It is the same for you. In Swahili, you may say, "….." when, if you think about it, or if you write it, it is actually two words: Swahili example:

4. Helping readers by making reading easy for them:

   - Symbols which are easy to distinguish, (Example: what if all of our letters were tall, like this: llttff!!lf)?

   - Shorter words, when possible

   - Giving the reader all the information he needs to be sure of the meaning of a word (including tone, the correct vowel, length, etc.)

     (Examples of contrastive vowel length and contrastive tone)

     What happens to your understanding of these sentences? What makes a difference in your understanding?

     - Mti ambao unafa

- Mti ambao unafaa

5. Helping readers understand what the writer wants to tell them (not just helping them sound out words so they can be pronounced). We use alphabets to help us do this. We choose letters to represent sounds. But a writing system is more than sounds. I could teach you the Spanish alphabet, A, B, C, Ch, D, E, F, G, H, I, J, K… but if you could repeat the sounds after me and sound them out in words, would you be reading? (Example: Quisiera leer mi idioma.) Did you understand? No, of course, you still need to get *meaning from words* in order to be truly reading.

Another goal we have for each language group is to record some of your stories before we leave. You might want to be thinking of an interesting one which your best storyteller can tell, later this week.

Appendix C

Lesson 2: Helping Readers with our Spelling Decisions

Lecture presented by Leila Schroeder in February 2007

Learning objectives: people will start thinking about readability/writeability issues in choosing graphemes:

- Visual discrimination of symbols

    What is difficult here for a new reader? Find the letter which is different in each row. (Chalkboard)

    αααoα (Can you easily find the o?)

    uunuu

    bdbbb

    pbbb

    uuûu

    tffff

- Simple sound-symbol connections (n vs. ng')

- Comprehension challenges (see examples below)

Lecture

In order to write what we say, we must

1. Break up the stream of speech : sinunuimkatemjinikwasababuhaufai
   Divide the stream into words: Sinunui mkate mjini kwa sababu…

2. Choose from the ancient Roman symbols available—the letters we need in order to write the sounds of our languages.
   These letters are a code developed by the Romans thousands of years ago, for their language, Latin.
   A B C D E F G H I J K L M N O P Q R S T U V W X Y Z.
   The Spanish needed some different letters, so their alphabet looked like this:
   A B C Ch D E F G H I J K L Ll M N Ñ OP Q R RR S T U V W X Y Z

3. Notice that there are very few digraphs in these alphabets, because there is almost a 1-to-1 correspondence between 1 symbol and 1 sound.

4. We can use those letters in any way we want, so they work for our language. Because all of you already know how to read, you can easily learn how to pronounce some new letters, or say them differently when you read a different language with them. For instance, with the Spanish language I could tell you that

- Ñ stands for "ny," and when you read these words, you will say the "ny" sound: say

  ni ña – 'girl'

  a ra ña – 'spider"

- Spanish uses a double l for a 'y' sound can you "read" these words?

  Llamo

  amarillo

  llora

- Now, here's another difference in the Spanish alphabet: they use "qu" instead of k, before e or i. Can you "read" these the Spanish way?

  Queso –'cheese' and

  Quiero – 'I love'?

  Now you all are ready to "read" Spanish. Yo quiero queso amarillo, para mi niña. (I want yellow cheese for my little girl)

You were not really reading, of course, because you couldn't understand what you said! But it does show you that once people have learned how to convert letters into sound, the rest is easy, especially with understanding!

5. Here is an important point about writing languages down: We help readers when we write down the way people talk, but in slow speech.
    If we wrote the way native English speakers speak English, we would lose many things which help a reader understand. For example, we often say: "Waijaduthat?" and "gimmithiapple." You must read them aloud, in order to figyuremaut—figure them out.

    We need writing systems which show slow speech, not words crowded together, so the reader always can recognize whole words.

    For example: na antû   and people   not nantû

                            ni wee   it is him   not nwee

6. Problem for readers: remembering a long series of sounds or syllables: okwifunyongora is harder than a word like rigori.

7. Readers have to tell the difference between all these letters. We may be able to help them by choosing symbols which are not too hard to tell apart.[1]
    α-o u-n b-d p-b u-û f-t nt-nf ng-nj ch-h bh-sh ng'-ng
    What are the challenges for the reader below? (Similar appearance of most of the graphemes)

    nduru

    nturu

    nguru

    njuru

---

[1] This section is taught later, in Lesson 4, in expanded form

8. The hardest letters for people to read are the digraphs (two or more letters for just one sound). Your language will need many of these, but whenever you can eliminate some of these and choose one letter from the Roman alphabet, I recommend it.

   For example, if you are writing bh and gh but you are not using the b and g at all, I recommend using the b and g. If you do this, you can simply tell the reader that in your language it makes a soft sound.

9. There may be a problem with this recommendation, if people really want their language to be written like Swahili, with gh meaning the softer sound of g. But after they try reading and writing with the longer spellings, they may complain that it takes too long to write them this way, and it is harder to read the words, especially when they are long.

What are challenges for writers?

- Long words
- Complicated spelling rules they must remember
- Words or morphemes that change spelling.

How do we help writers?

Give them spelling rules that are simple. In English, we always write some words the same, even though they are pronounced differently using phrase-level rules. For example,

We always say "The" as /thə/ before a consonant: "thə dog."

But before a word beginning with a vowel, the same word is pronounced /đi/. So we say "thi apple." Since it is actually the same word in the mind of the speaker, we make things simple for both the reader and the writer and spell "the" the same. A native speaker will almost always pronounce them according to the speech rules of his language. Let's try it with a native English speaker reading this list:

- the pencil
- the house
- the purple cow
- the orange

Did you hear the difference? It is natural to any native English speaker, even a child, though maybe not to you.

What are the challenges for new *readers* (review the list above)?

1. They're working on matching sounds with symbols

2. They're trying to move their eyes from left to right, remembering a series of sounds.

3. They put them together into wordsThey will sound out "Okusoma na Okwandika Ekikwaya" and they may say, "This is not the way we say it. We say, 'Okusomanokwandika Ekikwaya'." What will you tell them?

What about mature readers?

Our goal is also helping them, because they will be the people who can actually read something harder like the New Testament.

Mature readers recognize whole words, not just syllables or letters. If you write everything connected, with some letters missing, as in rapid speech, these more advanced readers may come to you complaining that it's hard for them to read quickly and with understanding. They are forced to read slowly and out loud, in order to understand. They will ask you to make separate words of these things, so they can "see the meaning."

Appendix D

Lesson 3: Consonant Symbol Choices

Lecture presented by Leila Schroeder in February 2007

Learning objective: People will demonstrate an increased awareness of the rationale behind grapheme choices. They will make decisions which reflect not only their attitudes toward other languages and the sounds of their language, but the needs of beginning and advanced readers.

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z		

Lecture

1. ABCs…on Chalkboard: This is the ancient Roman alphabet, with a few modifications. They had 26 sounds in their language and used one letter for each sound, a direct match.

2. What we try to do in any alphabet is to match up sounds and letters. You can use one symbol per sound if your language has just as many vowels and consonants as the Romans had. It doesn't! All of your languages have more than 26 sounds.

3. Are there any simple alphabet letters we don't need at all? (Erase Q, X, and V from big board…)

4. What sounds do you feel you have in your language which aren't represented here? Invite TAs from the four groups to come up and write on their chalkboards their unique letters!

- Ikizu has added two more letters for the vowels you need (Add ʉ and ɨ to the Ikizu chalkboard).

- Simbiti has a vowel diphthong: what is it? Ai.

- Most of you have added several letters for the prenasalized stops, because each one of them has its own sound: All of have **mb.** (Add mb to the chart.)

- Most of you have **nd.** (Add nd to the chart.)

- We have **ng**, etc…(Add ng to the chart.)

5. What are some other symbols that we (or some of us) must add to our charts? Ny, rr, ng'…

6. Are there still some Roman alphabet letters you are not using at all (i.e., **B** and **G**)? (Are **bh** and **gh** really needed, or can you use a simple **B** and a simple **G**? In answering these questions for yourselves, here are some things to think about:

Pro-B or G	Pro-Bh or Gh
Shorter words for reader	Show difference from Swahili
Easily recognizable letters	It's what people expect
Faster writing	Seeing the difference between bh and h, for example, can be hard for readers

Put on chalkboard, with some Kwaya writer (and another Kwaya reading them, top to bottom):

Kwaya:

Akagheera	abhoobha	okubhebhea
Akageera	abooba	okubebea

Simbiti:

Ibhibhurughusie	kebhoroghwe	ebhebhoroghwe
Ibiburugusi	ekeborogwe	ebeborogwe

You can all hear that the **g** and **b,** when they stand alone, are soft sounds, softer than what you hear in Swahili.

The **g** and **b** are your languages' shy, quiet sounds, which become brave only when a friend joins them (a nasal friend such as m or n)! They are actually the same sound, which you will see when you arrange the consonants on your charts at your tables: you will find that the prenasalized hard sounds only occur after an m or n, and the soft versions only occur when they are alone.

You can see that the roots are actually the same, in these examples. They represent the same sound, but harden after a nasal:

Simbiti

Orubharu	'ubavu'	sembaru	'mbavu'

Ikiizu

oroberi,	'kijiji'	zemberi	'vijiji'

Kwaya

Oribhubu	'firigisi'	gimamba	'hoja'

Write on the board and have people all read the following aloud. You can see as you read across, that native speakers cannot help themselves. The soft letter hardens after a nasal.

ighi (or ighi)	i**n**ghi
agha	a**n**gha
eghe	e**n**ghe
ubhu	u**m**bhu
ibhi	i**m**bhi
ebhe	e**m**bhe

Can you hear the *bravery* of those sounds when they join a nasal? They are the same sound, but they harden when they have their friend with them. It is impossible for you to pronounce the g with a soft sound after m or n. (They are the same sound and should be written with the same letter in both places.)

Conversely, when you speak Kikwaya, you will never say "Ariga" (with a hard sound for g). When you read Kikwaya it should be the same. Call for a volunteer Kwaya speaker. Instruct him to pronounce the bs and gs softly when he reads them:

- Ariga oriwo omunibhi umwi abheye na omgasi waye na abhana bhaye na omkasi waye owemirimu.

- Ariga oriwo omunibi umwi abeye na omgasi waye na abana baye na omkasi waye owemirimu.

If you were to instruct native Kwaya speakers that they are reading *their language*, not Swahili, they would probably very quickly learn to pronounce the words the way Kwaya speakers naturally speak, making them soft when they are alone, and making them hard after a nasal.

Should people write <**bhabhaabha**> 'fathers' or **babaaba**, then?

Conclusion: Many Bantu languages have a soft sound for **B** or **G,** when they are alone. The sound becomes harder when a nasal is with it.

- Native speakers of your language will very quickly learn to say the soft sound of B and G when they read aloud, because it is natural for them to do this. But—

- If people really don't like writing just b or g, you may want to choose the digraphs, or a symbol like β or Bh. We don't want people rejecting their alphabet!!

Repeat choices:

Pro-B or G	Pro-Bh or Gh
Shorter words for reader	how different from Swahili
Easily recognizable letters	It's what people probably expect
Faster writing	Seeing the difference between bh and h, for example, can be hard

So this is your choice: B, β, or Bh,   G, or Gh

Appendix E

Lesson 4: Morphemes and Words

As we have said, the challenges for readers and writers are:

1. long words: okwifunyongora

2. spelling rules: Only write a vowel which sounds long as long when it …

3. words and meaningful parts of words which change spelling (Example: ku-pasua, ku-tengeneza, kulala, kwenda)

Some of the things we do to make reading easier:

1. Break up the stream of speech (Example on chalkboard):
   Sinumuimkatemjinikwasababuhaufai

   We don't write it like we say it, because it is too hard for the person looking rather than listening, to make sense of it!).
   Instead, we break up that stream of speech into separate words, even though when we speak it usually all sounds like one big loooong stream of speech. We make this look like: Sinunui mkate mjini kwa sababu haufai.

2. We try to make these separate words look the same every time the reader sees them. In English words, for example, we have question words. Almost all of them start with the letters <wh>. This helps people recognize them as soon as they see them written. We have **why, what, when, who, and what.** The <**wh**> beginning immediately signals a reader that this is a question word! So a simple spelling rule helps readers get meaning quickly when they read.

3. We try to keep the spelling of words the same, even though their sound may change when they are spoken fast in a sentence. For example, Americans may say, "Waijaduthat?" (Write on chalkboard). We write it like this: Why did you do that? But when we are really good readers reading aloud, we still say: "Waijaduthat?," not "Why did you do that?" And "gimmethupencil," not "Give me the pencil." So, we spell words the way we speak when we're talking slowly. We need writing systems which show slow speech, not words crowded together, so the reader always can recognize whole words.

   Tharaka example: na or ni? In Kitharaka, *na* means 'and'. *Ni* means 'it is'.

   The Tharaka people thought that if they could write just the way they speak, they would combine words like this: na antü would be written <nantü> 'and people'.

So:	na antû	'and people'	not nantû
	ni wee	'it is him'	not nwee

   If both *na* and *ni* are written attached to the word which follows them, with the vowel dropped, the reader will not be able to tell whether the little word was *na* or *ni.* Also, if they are written together with the word which follows them, you are

combining a verb with a noun, which are two distinct parts of speech. The reader can't easily tell whether the writer meant to say "it is people" or "and people," because both would be written <nantû>. Can you see the solution? Writing them both fully, as separate words. Will this be a problem for the reader? Only for a beginner. The beginner will quickly learn to get meaning and naturalness when reading fast, because he/she will recognize the meaning and be used to speaking his own language naturally.

Try this with an example from some language spoken at the workshop.

Appendix F

Lesson 5: For TAs: Spelling Principles

1. Small words, especially grammatical ones, should always be written the same, when possible (i.e., when native speakers aren't acutely aware of a phonemic distinction). Write them disjunctively when possible. If they're attached phonologically, try a hyphen with people.

2. Apostrophes should represent **one** thing when possible. English uses them for vowel elision and also for possessives, which confuses highly literate people all the time. **It's** or **its** (apostrophe problem).

   - Multiple uses for it

   - Its position, detached from the line of letters. Hyphen is sitting down with the vowels…

   Since we're stuck (sociolinguistically) with ng' in Eastern Africa, we should avoid other apostrophes dangling in the air, and let them just be part of this trigraph.

3. Lake Region vowels—the devoiced Vs in 2nd syllable of CVCV roots—due to their lack of stress, they may lose some of their contrastiveness, like English does in unstressed syllables, where everything becomes [ə].

4. Any questions about distinguishing palatalized consonants from CV sequences? (in Linguistic Features, Consonants, in Part 1, chapter 2) Consider phonotactic rules of the language. Strive for spelling consistency and consider morpheme preservation issues. Look at the Tharaka examples showing what often happens to Vs after palatals. Tharaka Vowel "Coalescence" chart. Writing of the palatal sound as a <y> versus as an <i> or <î> may depend on what the original V was underlyingly, more than upon some phonological rule. Preserving a distinction which would be otherwise lost, if both Vs, for example, in a 7-vowel system.

5. Palatal Nasals, p. 31-34. If your language has palatal, palatalized, and palatalized palatal nasals, they probably need to be distinguished in the writing system (especially if they occur frequently or their spelling affects people's recognition of affixes within words).

6. Labialized consonants, same. [kwe] versus semivowel following a V, like [kuwe]. See p. 35-40. A phonological distinction phonologically may signal a crucial distinction morphologically. Listen for tone changes, and ask your linguistic consultant! She knows the morphemic clues. Here's a phonological clue:

   If the vowel after the possible semivowel sounds long, then it really is a semivowel.

7. See Part 1 whenever you have questions, but most symbols will be obvious and intuitive.

Appendix G

Lesson 6: Goals and Products of Workshop 2

Instructional Objectives:

- Participants will be able to list their accomplishments from the last workshop
- They will be able to explain the rationale behind their spelling decisions from the last workshop
- Participants will see the value of studying their morphology in making spelling rules
- They will be able to list what they expect to produce by the end of this workshop

Let's review what we learned and accomplished last time, and then we will discuss the plans for the next three weeks.

What we did last time:

- Week 1: Vowels: **Irabu**
  - Vowel inventory (and Chart of Vowels filled, with example words in Part 2, Appendix L, Orthography Workbook Form)
  - Vowel combinations (chart filled; see Part 2, Appendix L, Orthography Workbook Form)
  - Long vowels where do they occur? (occurrence in disyllabic word roots; see Part 2, Appendix L, Orthography Workbook Form)
  - We studied two kinds of words. What are they? Nouns *nomino* and verbs *vitenzi.*
  - Noun roots are identified (filling noun class charts, just like Swahili has). Put up a sample noun class chart for one language group.
  - Check spellings of wordlists, using cards and giving ourselves spelling rules based upon certain rules for Bantu languages. Can you remember what they were, for all of us? When do you automatically write a vowel as short? (Before prenasalized stops: <nd, nj, ng, nz, mb …>). Does this apply to <ny> and <ng'>? Why?
- Week 2: Consonants: **Konsonanti**
  - We said that it is good to use only one symbol for each sound in your language, but that sometimes this is impossible, both for writing and for reading. For this reason most of you chose to write your soft consonant sounds with one letter, such as <b> or <g>.
  - We said that some sounds, actual letters in your alphabet, can harden or soften, even though they are the same phoneme or letter in your alphabet. This is what happens with
    - o Root-initial consonants
    - o Root-medial consonants
  - We learned the possible locations of each consonant we noticed, after studying the consonants in different positions in a word, that some change in pronunciation. For instance, you never find a [β] after an /m/. It always

becomes a [b] sound in your languages, after an <m>. Example: You can
see that the roots are actually the same, in these examples. They represent
the same sound, but harden after a nasal:

Simbiti: **orubaru** *'ubavu'*   **sembaru** *'mbavu'*
Ikiizu: **oroberi** *'kijiji'*   **zemberi** *'vijiji'*
Kwaya: **oribhubu** *'firigisi'*   **gimbugu** *'firigisi'* pl.

- Native speakers don't usually notice this change. You may be aware of it
  because your [β] is different from Swahili's [b]. Readers of Kikwaya or
  Kingoreme, though, will know which way to pronounce it correctly for
  Kikwaya or Kingoreme.
- Choose key words and pictures for alphabet charts
- Revised spellings of your wordlists
- We included w and y in all of your alphabet charts—but now we will look at
  them much more closely! I will explain this later.

- Week 3
  - Tone study begun
  - Noun class charts partially filled (show Ngoreme chart)
  - Alphabet charts produced
  - Short stories recorded, with transcription and word-by-word Swahili gloss
    and phrase-by-phrase free translation

These will be your biggest challenges:

Our goals for this workshop:

- Find out people's response to the wordlists
- Fix up our alphabet charts
- Discover some spelling rules and document them
- Together we will need to write down grammatical words and particles and list them
  on charts. This time you will make your own charts, as well.
- Test the readability, comprehensibility, and acceptability of certain spelling choices
- Stories: Correct some errors in accurately transcribing what storytellers said. If we
  have time, edit the spelling and punctuation of one story, to be printed out at the
  end of the workshop
- We will pay a lot of attention to the places where meaningful parts of words come
  together, because they make a big difference to writers and readers. The linguistic
  consultant will be helping us identify these parts.

  The pastor walks into his office and finds three people in it. He finds a note from
  the church secretary on his desk. What is the problem with the message he reads?

  - *Wanazima vitabu.* The secretary intended to say, "They are borrowing…"
    What do you have to do to make the meaning clear?
  - *Mbwa wake anacha nyumba.* Can you just write one <a> in these words?
    - *Wanaambia*: We can give it a completely different meaning by changing

only one tiny vowel: now it says *wanaimbia*.

- o In Kiikizu: *ikiriro* (kilio) *ikiriiro* (mboga)
- o In Kwaya: *risina* (bulb) *risiina* (name)

Is some information missing? (And what happens to your understanding of these sentences when just one letter is added?)

What makes a difference in your understanding?

Did one little letter make a big difference in your understanding of a whole word?

Break this word into parts: *wa-na-ambia*. (Now substitute *wanaimbia*.)

When *viambishi* come together, we often have spelling challenges. Karen will explain these in detail, as we look at nouns first, and later as we study verbs.

- – You will see during this workshop that your vowels sometimes change and become consonants (w or y) in some situations. Look at Noun class 1 prefix for kiNgoreme. For example, in Kingoreme, what vowel here became a w?

Noun class 1 prefix *omu-*

prefix	+	vowel	becomes	orthographic result	example word	Gloss
omu*	+	a	▶	omwa	omwame	'bwana'
omu	+	e	▶	omwe	omwenekere	'mwenyeji'

*Danny Foster adds (pc), "We don't know if [the] underlying [form] is *omo* or *omu*. I've chosen the high vowel. The prefix vowel is subject to vowel harmony rules when the root begins with a consonant."

Literacy Consultant's job:

- • Help find solutions to reading-spelling challenges
- • Help document what we learn
- • Make good alphabet charts and other literacy materials

Linguistic Consultant's job:

- • Helping you discover the grammar of your language
- • Helping you discover as many of your grammatical words as possible
- • Uncovering the rules you aren't even aware of, which make things work the way they do in your language (like we did last time, in looking at your noun class prefixes) and what happens to them to make them take different forms… Participants' job:
- • Give us examples from your language when we ask for them, as you did before
- • Try reading things different ways (i.e., with a variety of spellings) and give me feedback on which way makes understanding and spelling easiest for you
- • You can make charts of how your grammar works after we have made the discoveries. The more you learn about your grammar, the better you will spell and

read your language and the better you will explain spelling rules to others! We will all make one big chart with each language group, to start with. Later, each of you can fill in a chart with the same information—not yet.

Review the following orthography principles, if needed

Remember, your goals in developing your writing system:

1. Helping writers by making spelling easy, with simple rules

2. Helping new readers by using one symbol for each sound in your language whenever possible

3. Helping new readers by breaking up the sounds into words and sentences. Example: sikununuaviazimjinikwasababunilikosawakati Exercise: Break into words— Sikununua viazi mjini kwa sababu nilikosa wakati.

4. Slow speech. When we talk to our friends, we usually drop some of the sounds in words, knowing they will understand us. In English, when we talk fast, we say, "G'bye." But if we speak slowly, we are actually saying "Good Bye." That is the way we need to write it: "Good Bye." We native English speakers say, "Dijuseeimatth store?" But when we write, we break these sounds of fast speech into slow speech, and really help the reader and writer to understand: "Did you see him at…" It is the same for you. What does this (write it) mean to you? <Shkm>. Can you tell this is "shikamoo"? Readers, unlike listeners, need all of the clues we use in slow speech to help them understand.

5. Helping readers by making reading easy for them:

   • symbols which are easy to distinguish (Example: what if all of our letters were tall, like this: llttff!!ɟf?)

   • shorter words, when possible

   • giving the reader all the information he needs to be sure of the meaning of a word (including tone, the correct vowel, length, etc.)

     (Examples of contrastive vowel length and contrastive tone?)

   • Helping readers understand what the writer wants to tell them (not just helping them sound out words so they can be pronounced) (Example: *Quisiera leer mi idioma.*) ¨Did you understand? No, of course, you still need to get meaning from words in order to be truly reading. You're not just making sounds. Have you noticed that a really advanced reader can read without making a sound? He recognized several letters and sometimes whole words at once.

6. Alphabets are just a tool, not the ten commandments from God, handed down in stone in English! The letters do not even represent the same sounds for all languages. You get to decide what sound each letter stands for, for your language. Example: In Swahili, <ch> is the [ch] sound. In French, <ch> is the sh[š] sound. *Chien* – dog. In Czeckoslovakian, they write it with a <cz>. The [ch] sound in French is written this way: <tch> . In Swahili, [y] is written <y>, but in Spanish, it is usually written <ll>, as in *Me llamo.*

Of course, we also want to make it easy for people who already read Kiswahili to learn to read their mother-tongue, so we keep many Swahili symbols and sounds the same in the mother tongue.

Tharaka example for production of Semi-Vowels when vowels converge at morpheme breaks

It was noted before that w and y are called 'semi-vowels'. They have an influence upon the vowels that follow them by making them sound long, when in fact there is only one short vowel present. For example, *watho* 'law' will sound as if a long aa sound is present. But we know that the root of the word is *-atha*, to which is added the (abstract) noun class prefix *û*, which changes to w before a. Since the underlying form only has one vowel, only one vowel will be written. This is still pronounced long.

The same applies to the other semi-vowel y. When this occurs before a single vowel, that vowel will sound as if it is long. For example, *yake* 'his'. This is made up of *î* + *-ake*. It sounds long, but there is only one a vowel present in the underlying form. Here are some other examples:–

û + onokio	**= wonokio**	**î + akwa**	**= yakwa**
û + ega	**= wega**	**î + enu**	**= yenu**

When two identical vowels follow a semi-vowel, then both vowels will be written. This is a long syllable and will sound long with or without the presence of the semi-vowel. For example, 'you built' (distant past) is *waakire*. Here the different parts making up the word are: û + a + aka + ire. Here we see that one **a** vowel is the past tense marker, and the other a is the first letter of the word 'build'. Other examples are:–

û +a + anjîîria	= waanjîîria	î + a + amba	= yaamba
û + a + athaga	= waathaga	î + a + akire	= yaakire

Also note that two vowels that join together to form one long vowel will also both be written after a semi-vowel

wa + ûgu	= woogu
ya + ûgu	= yoogu

So basically, the rule is a simple one: where there is a SV syllable which only has one vowel in the underlying structure, then only one vowel will be written, but where there are two vowels in the underlying form, then two vowels will be written. In pronunciation it is possible that both will sound the same.

Appendix H

Lesson 7: Word Breaks

Instructional objective: Participants will make good decisions as to word-break rules, regarding clitics and other grammatical words or particles which may be affected by phrase-level rules in their pronunciation. They will understand the options presented to them, and their choices will reflect their perception of their language as well as readability and writeability issues.

Materials:

- chart listing word break principles, in language of wider communication (the chart is partially covered, at the beginning, so that all that shows is the title: What is a Word?

- flipchart with paper, chalkboard, or whiteboard (for examples)

> ### What is a word?
>
> 1. It has meaning all by itself.
> 2. It can move around and be put in different places in a sentence
> 3. Other words can come *in between* words

When people first begin to read, they sound out, letter-by-letter, syllable, without always understanding what they read. Why? Because they are working very hard just to make sounds which match letters.

A beginning reader looks at <**a-bhe-ju-ku-ru**> 'grandchildren' in Kikwaya, and by the time he reaches the end of the word, he may not even remember what it all said, to put the syllables together to make words. He's just seeing letters and making sounds. That is normal.

But if a reader gets lots of practice, he can in a few months start sounding out words faster and putting those sounds together into words. Later, a really fluent reader, such as the people in this room, can just glance and see a whole word at once. This is especially true for words which are frequently used. Shorter words, such as *nyumba, mtu, wa, kubwa, mti, ni, nani, je,* etc., are used so often that a practiced reader recognizes them as a *whole word.* He is not sounding out **na-ni.**

Let's think first about what *makes* a word. A word:

1. It has *meaning* all by itself, though some words have more distinct meaning than others. Nouns, like **nyumba** and **watoto**, have lots of meaning on their own, and when we think of them we have a picture in our minds. That's why we used nouns for your alphabet charts. (Talk about adjectives and numbers and the way they can be picturable and isolatable.)

2. Most adjectives and pronouns also have distinct meaning, and you can still picture them a bit: simple pronouns like **wewe**, number words like **kumi**, and even roots of verbs, like -**lala** or -**chagua.**

3. Most Bantu words, especially verbs, contain grammatical indicators which must be there, in order to make sense. We will look at them next week. That means that you have a *part* of a word which carries most of the "wordness." In order to really make sense in a Bantu language, you have to add something to the verbal root, making something (mzizi) such as <**a-na-cheka**>. Just briefly, what information do you get from this verb and its grammatical attachments? Many more things can be added to a Bantu verb, giving much more information, but this much is essential. This much grammatical information is usually essential to a Bantu verb in order for it to make sense, so it is all *part* of one word (an affix, not a word).

4. Can the **a-** in **anacheka** communicate any meaning by itself? No, it needs the rest of the word in order to make any sense at all.

5. A word can *move around* in a sentence. For example, take the word **nyumba.** You can put it at the front of a sentence: **Nyumba yake ni kubwa**. You can put it after a verbal phrase: **Wanapenda nyumba ya mtu huyu**. The **a** in *anacheka* cannot move from that exact place in the verb. It must be right there at the front, telling the listener or the reader *who* is doing the laughing, and it has no meaning on its own. As you probably know by now, **a-** in **anacheka** is an affix, not a word.

6. Words can be written separately when other words can come in between them. For example, in Swahili, *ya* in a sentence like *Nyumba ya mdogo wangu* 'the house of my younger sibling', is it possible to put another word between **ya** and **mdogo**? Yes, you can say **nyumba ya huyu mdogo wango.** So it should be written separately because it is possible.

7. Sometimes you have very little words in Bantu languages; they carry their own little bit of meaning, but people always say them with a noun. The following is a locative example from Nyiha. People recognize them as separate words because the nouns *nyumba* and *mbwa* can stand all by themselves. It's easy to recognize **mbwa** and **nyumba** by themselves. But a spelling problem always arises when the noun starts with a vowel, and you put the locative before it! Then you get spellings like the noun **ikwi** becoming **pikwi, kwikwi,** and **mwikwi.** Two words have been written as one, because the vowels at the end of the locative changed and so did the vowels at the beginning of the noun <**ikwi**>. Now the noun is merged with another word and it's hard for the reader to recognize the noun.

8. Locatives and associatives for your languages will probably turn out to be small, separate words. Other words can intervene between them, they have meaning of their own, and they're movable.

What should we do to resolve the writing problem above? The Niha have choices:

1. Write the locative the same everywhere and write the noun the same everywhere. Then you have <**pa ikwi**>, <**ku ikwi**>, **and** <**mu ikwi**>. Native speakers will eventually learn to say the words naturally in fluent reading: /**pikwi, kwikwi, mwikwi**/, because this is how they talk.

2. Write the locative attached to the noun it describes, with a hyphen: **pa-ikwi, ku-ikwi,** etc.

3. Write the locative combined with the vowel prefix from the noun, like this: **pi kwi, kwi kwi, mwi kwi.** This makes the noun look different, harder to recognize now because its preprefix has been removed. It also makes the locative harder for the reader to recognize because the <**a**>, <**u**>, and <**i**> are no longer there as a clue.

4. Write all the changes as people *say* them, with the locative attached to the noun it preceded: <**piwki**>. It makes the noun harder for people to read, because it is no longer standing alone, and it always looks different because it has these locatives stuck onto it in such a way that it's hard to cut apart the locative from the noun— but it's written the way it sounds, making it easy for a new reader to sound out, letter by letter. It will slow reading for a fluent reader.

5. The first two options are the best. Today we will be looking at locatives, associatives, and possessives and tackling these problems for writing. (You actually have the same issue with your prefixes on your nouns.)

Any questions?

Appendix I

Lesson 8: Story Editing

Objectives:

- People remember their writing rules.
- They can apply them to editing a story.
- They test their rules by reading the story aloud.
- They note any unnoticed linguistic elements which need to be documented, while editing.

Materials:

- A handout chart for each group to put up, like the Ngoreme one at bottom.

Activities:

- Review the concepts we taught earlier, re: what makes a word and the reason for word breaks.
- Look at their charts to see what spelling rules they have established already.

List them, for:

- **Locatives**: Example from Ngoreme:

ko-mugabo	mo-mugabo
ko-moona	mo-moona
ko-monto	mo-monto

- **Associatives**

Table 1

egikoba	kyo	moona	kya	baana	kye	nyumba
ebikoba	byo	moona	bya	baana	bye	nyumba
engoha	yo	moona	ya	baana	ye	nyumba

- Copulas—do they want to write the copula as attached, but to write the *na* conjunctive separately?

  Focus marker /n-/ is written adjoiningly. There seems to never be an underlying vowel with it. Since they have the copula *n-* which is very snugly attached to the following noun or adjective, do they always want to write it adjoiningly? Or do they want to write it with an apostrophe after it? It will look just like the conjunction if they write them both at the front of a word (usually a noun) with no vowel after either one.

- Augments/diminutives—nothing unusual
- Negatives—written as part of the verbal phrase (*tu*)

140

- Other, including tone: extensions (causative and passive are interesting—they are just *i* and *u,* respectively)
- Vowel length? Ask us when problems arise, and we'll talk! At issue: simplicity of spelling rules versus morphology. Kwaya example: woone.

Table 2: Ngoreme example

-rogota		-agura	-engeri	-ikara	-ogu	-u
		'-tambaa'	'-ongezea'	'-kaa'	'-sikia'	
twakarogota	'twaokota'	twakaagura	twakeengeri	twagaikara	twakoogu	
turarogota	'tunaokota'	turaagura	tureengeri	turaikara	turoogu	
nturogotire	'tumeokota'	ntwagurre	ntwengeriiri	ntwikaire	ntoogure	
ntwarogotire	'tuliokota'	ntwaagurre	ntweengeriiri	ntwaikaire	ntoogure	
tukarogota	'tukaokota'	tukaagura	tukengeri	tugaikara	tukoogu	
tokurogota	'tukiokota'					

n-tu-rogot-ire

n-tu-ogu-ire = > ntoogure

n-tu-a-rogot-ire = > ntwarogotire

n-tu-a-ogu-ire = > ntoogure

<div>

Sheria za Kuandika Kingoreme

1. Mahali: (ko-mugabo, mo-mugabo, mo-moona, ko-moona, ko-monto)
2. –a unganifu: (wo bo go gyo) **moona** and (wa, ba, gwa, gya) **baana** and (we, be, gwe, gye) **nyumba**
3. Kitenzi shiritisho (ni): (before C? before V?)
4. Negatives: ??
5. Vowel length? Write short before prenasalized consonants and after semivowels. Violate this rule only: "When you must show something which makes a grammatical difference in a verb…"

</div>

Appendix J

Noun-Class Prefixes for Proto-Bantu (Gardner 2005:54)

Class	Nominal Prefix (NPx)	Pronominal Prefix (PPx)	Enumerative Prefix (EPx)
1 sg	mʊ-	jʊ-	ʊ-
2 pl	ba-	bá-	bá-
3 sg	mʊ-	gʊ-	ʊ-
4 pl	mɪ-	gɪ-	ɪ-
5 sg	i-	lɪ-	lɪ-
6 pl	ma-	gá-	á-
7 sg	kɪ-	kɪ-	kɪ-
8 pl	bi-	bi-	bi-
9 sg	N-	jɪ-	ɪ-
10 pl	N-	jɪ-	í-
11 sg	lʊ-	lʊ-	lʊ-
12 sg	ka-	ká-	ká-
13 pl	tʊ-	tʊ-	tʊ-
14 sg/pl	bʊ-	bʊ-	bʊ-
15 sg	kʊ-	kʊ-	kʊ-
16 locative	pa-	pá-	pá-
17 locative	kʊ-	kʊ-	kʊ-
18 locative	mʊ-	mʊ-	mʊ-
19	pi-	pí-	pí-
20	gʊ-	gʊ-	gʊ-
21	gi-	gí-	gí-
22	ga-	gá-	gá-
23 locative	ɪ-	ɪ-	ɪ-

These are a synthesis of Schadeberg 2003:149 and Katamba 2003:104.

Typical singular-plural pairings of Bantu noun classes

Class #	Singular prefix	Class #	Plural prefix
1	mʊ-	2	ba-
1a	Ø	2/2a	Ø
3	mʊ-	4	mɪ
5	i-	6	ma-
7	kɪ-	8	bɪ-
9	N-	10	N-
11	lʊ	10	N-
12	ka-	13	tʊ-
14	bʊ-	6	ma-
15	kʊ-	6	ma-
16	pa-	--	--
17	kʊ-	--	--
18	mʊ-	--	--
19	pi-	13	--
20	gʊ	22	ga-
21	gɪ-	--	--
23	ɪ	--	--

The chart below (Katamba 2003), in addition to the simplified Proto-Bantu chart of noun class prefixes, may help the linguist recognize the noun class prefixes for a given language.

Semantic Content of Bantu Noun Classes

Class	Semantic content
1	Human beings
1a	Proper names
	Kinship terms
	Personifications
2	Regular plural of class 1
2b	Regular plural of class 1a
3	Natural phenomena
	Body parts
	Plants
	Animals
4	Regular plural of class 3
5	Natural phenomena
	Animals
	Body parts
	Collective nouns
	Undesirable people
	Augmentatives
	Derogatives
6	Regular plural of classes 5 and 14
	Mass terms and liquids
	Time references
	Mannerisms
	Modes of action
7	Body parts
	Tools, instruments, and utensils
	Animals and insects
	Languages
	Diseases
	Outstanding people
	Amelioratives
	Derogatives
	Augmentatives
	Curtatives (shortness, stoutness)
	Mannerisms

Class	Semantic content
8	Regular plural of class 7
9	Animals
	People
	Body parts
	Tools, instruments, and household effects
10	Regular plural of class 9
11	Long, thin entitities
	Languages
	Body parts
	Natural phenomena
	Implements, utensils, and other artifacts ("artefacts")
12	Augmentatives
	Derogatives
	Diminutives
	Amelioratives
13	Regular plural of class 12
14	Abstracts
	Collectives
15	Infinitives, a few paired body parts
16	Location terms
17	Location terms
18	Location terms
19	Diminutives
20	Derogatives
	Augmentatives
	Diminutives
	Amelioratives
	Mannerisms
21	Augmentatives
	Derogatives
22	Plural of class 20
23	Location terms

Appendix K

Duruma Noun-Class Morphophonology

by
Participants of Duruma Writers' Workshop
together with
Constance Kutsch-Lojenga
at Kinango, Kenya

(updated version: November 1994)

phoneme		grapheme	sample word	gloss
class 1: mu-				
mu + i	→	mwi	mwivi	*thief*
mu + e	→	mwe	mwelevu	*clever person*
mu + a	→	mwa	mwananche	*child*
mu + o	→	mo	mokoli	*saviour*
mu + u	→	mu	mumbi	*creator*
class 2: a-				
a + i	→	ai	aivi	*thieves*
a + e	→	ae	aelevu	*clever people*
a + a	→	a	anache	*children*
a + o	→	ao	aokoli	*saviours*
a + u	→	au	aumbi	*creators*
class 3: mu-				
mu + i	→	mwi	mwiri	*body*
mu + e	→	mwe	mwezi	*month*
mu + a	→	mwa	mwadine	*tree, sp.*
mu + o	→	mo	moho	*fire*
mu + u	→	mu	muho	*river*
class 4: mi-				
mi + i	→	mi	miri	*bodies*
mi + e	→	mye	myezi	*months*
mi + a	→	mya	myadine	*trees, sp.*
mi + o	→	myo	myoho	*fires*
mi + u	→	myu	myuho	*rivers*
class 5: ø-				
ø + i	→	i	imbi	*wave*
ø + e	→	e	embe	*mango*
ø + a	→	a	aphwa	*fruit of eggplant*
ø + o	→	o	oza	
ø + u	→	u	uwa	*flower*

phoneme		grapheme	sample word	gloss
class 6: ma-				
ma + i	→	mai	maimbi	*waves*
ma + e	→	mae	maembe	*mangos*
ma + a	→	ma	maphwa	*fruits of egg pl.*
ma + o	→	mao	maoza	
ma + u	→	mau	mauwa	*flowers*
class 7: chi-				
chi + i	→	chi	chinu	*mortar*
chi + e	→	che	chera	*tool*
chi + a	→	cha	chango	*hanging basket*
chi + o	→	cho	chombo	*tr. house length*
chi + u	→	chu	chula	*frog*
class 8: vi-				
vi + i	→	vi	vinu	*mortars*
vi + e	→	vye	vyera	*tools*
vi + a	→	vya	vyango	*hanging baskets*
vi + o	→	vyo	vyombo	*tr. house lengths*
vi + u	→	vyu	vyula	*frogs*
class 11: lu-				
lu + i	→	lwi	lwingu	*spleen*
lu + e	→	lwe	lwembe	*blade*
lu + a	→	lwa	lwala	*grinding stone*
lu + o	→	lo	lomba	*sharp point*
lu + u	→	lu	lungo	*winnowing tray*
class 14: u-				
u + i	→	wi	wira	*song*
u + e	→	we	wembe	*razor blade*
u + a	→	wa	wari	*food?*
u + o	→	wo	wongo	*brain*
u + u	→	u	unga	*flour*
class 15: ku-				
ku + i	→	kwi	kwimba	*to sing*
ku + e	→	kwe	kwenda	*to go*
ku + a	→	kwa	kwamba	*to say*
ku + o	→	ko	kokola	*to save*
ku + u	→	ku	kuza	*to ask*

DURUMA Noun Classes 9 and 10

		phoneme	grapheme	sample word	gloss
N + b	→	mb	mb	mbuzi	goat
N + v	→	Mv	mv	mvula	rain
N + d	→	nd	nd	ndenge	he-goat
N + dz	→	ndz	ndz	ndzovu	elephant
N + j	→	ɲj	nj	njira	path
N + g	→	ŋg	ng	ngoma	drum
N + gb	→	ŋmgb	ngw	ngwena	crocodile
N + z	→	nz	nz	nzugu	groundnuts
N + l	→	nd	nd	ndimi	tongues
N + p	→	pH	p'	p'anya	rat
N + t	→	tH	t'	t'iya	giraffe
N + ts	→	tsH	ts'	ts'ui	leopard
N + tʃ	→	tʃH	ch'	ch'ingo	hide, skin
N + k	→	kH	k'	k'uk'u	chicken
N + kp	→	kpH	kp'	kp'eza	octopus
N + ß	→	pH	p'	p'era	rhinoceros
N + m	→	m	m	manga	cassava
N + n	→	n	n	nazi	coconut
N + ɲ	→	ɲ	ny	nyenyezi	star
N + ŋ	→	ŋ	ng'	ng'ombe	cow
N + f	→	f	f	figo	kidney
N + s	→	s	s	saa	watch, clock
N + ʃ	→	ʃ	sh	shati	shirt
N + h	→	h	h	hamira	yeast
N + i	→	ɲi	nyi	nyingu	spleens
N + e	→	ɲe	nye	nyembe	razor blades
N + a	→	ɲa	nya	nyala	grinding stones

Orthography Workbook Forms

Acknowledgments

Dr. Constance Kutsch-Lojenga was my original advisor for documentation of the phonology of each language to be studied. After that important planning stage, Karen van Otterloo added to the charts we designed, providing many more, especially for studying and recording morphology. Finally, I have attempted to include a simple means of recording spelling decisions and their rationale in the documentation process, with inspiration from Danny Foster, Training Coordinator for SIL, Uganda-Tanzania Branch. I'm grateful to all three of you!

Leila Schroeder

Contents

Orthography Workbook Form

Phonemic Data

Name of language _____

Please read chapters 1 and 2 of Part 1 before beginning the process of orthography development. Chapter 3 will guide you in choosing of symbols and word breaks.

Vowels

**Nine-vowel system _____   Seven-vowel system _____   Five-vowel system _____
Other _____**

List symbol choices below. Add rationale for any non-Roman symbols, at bottom of the page.

Table 1: Vowel Inventory

IPA	Symbol	Sample Word (Disyllabic root with 2 identical vowels)	Gloss
		Sample	

Table 2: V1 - V2 combinations chart

This chart shows the possible vowel combinations in disyllabic word roots. List all the vowels of the language across the top and side. The vowels listed on the left side are those which come first; those along the top come second.

Write the first five or seven vowels down the left column, and the same vowels across the top. The sample words written in each box will have a Swahili gloss. They should all have disyllabic roots, with series of 2 vowels. The first table gives a partial example from a hypothetical language.

Example Table

V2 / V1	a	e	i	o	u
a	litata fimbo	kitate kikapu	kitati kiazi	kitato kiatu	kitatu kifaru
e	kiteta hotuba	kitete ndege(namna)	kiteti samaki (namna)	luteto mti (namna)	kitetu mto
i					
o					
u					

V2 / V1						
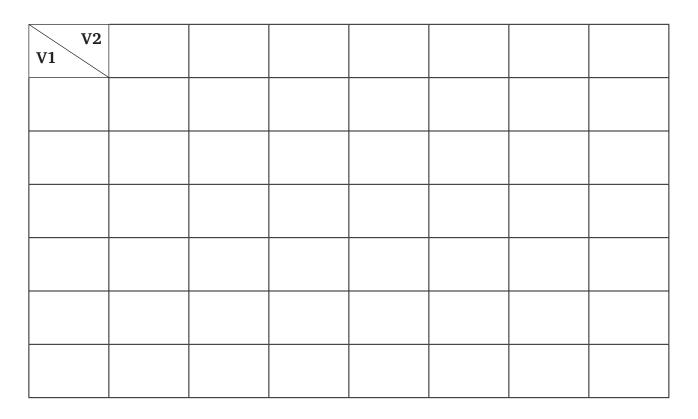
Table 3: Long Vowel occurrence in disyllabic word roots

Write long vowels in first column. Give one example for each vowel of the language, with CV following the long vowel syllable and another example in which an NCV syllable follows. Follow the orthography rules that have been decided upon for vowel length. Make the two examples as similar as possible, e.g., i… omu-siiya… omu-simbe.

Vowels	-CVVCV	-CVVNCV

Table 4: Vowel Length

Vowel length: Is it contrastive (see Part 1, chapter 2, Vowel Length)?

Yes _____   No _____

	Short		Long	
Vowels	**Example**	**Gloss**	**Example**	**Gloss**

Table 5: Vowel Harmony

Vowel harmony: Yes _____   No _____

Give examples:

_____

_____

_____

_____

Look for examples of a verb in the given verbal extension, with i, e, etc., in the stem, to see what other vowels appear in the following syllables

	i	e				
kutenda	*kuimba*	*kupenda*				
kutendea	*kuimbia*	*kupendea*				
kutendesha	*kuimbisha*	*kupendeza*				
kutendwa	*kuimbwa*	*kupendwa*				
kutendeka						
kutendana						
kutendua						
kutenduka						
kutendama						
atendire						

Consonants

(see Part 1, chapter 2, Consonants)

Consonants of _____

Consonants: Root-initial

Write the grapheme chosen for your language in the appropriate box of the chart below.

Manner of articulation	Voicing/ nasalization	Labial/ labio-dental	alveolar	Palatal/ alveo-palatal	velar/ glottal
Implosives	voiceless				
	prenasalized				
	voiced				
	prenasalized				
Stops and affricates	voiceless				
	prenasalised				
	voiced				
	prenasalised				
Fricatives	voiceless				
	prenasalised				
	voiced				
	prenasalised				
Sonorants	nasal				
	oral				

Consonants which can be labialized:

_____

_____

Consonants which can be palatalized:

_____

_____

Limited distribution of vowels following these consonants, CVCV:

_____

_____

Consonants: Root-medial

Manner of articulation	Voicing/ nasalization	Labial/ labio- dental	alveolar	Palatal/ alveo- palatal	velar/ glottal
Implosives	voiceless				
	prenasalized				
	voiced				
	prenasalized				
Stops	voiceless				
	prenasalised				
	voiced				
	prenasalised				
Fricatives	voiceless				
	prenasalised				
	voiced				
	Prenasalised				
Sonorants	nasal				
	oral				

List any consonants which only occur root-medially, with examples:

_____

_____

Describe any complementary distribution of consonants in the two positions:

_____

_____

_____

List of examples for each consonant phoneme

Write the letter on the left, and on the right, a sample word with underlined syllable in focus. Each syllable has the vowel /a/

	Example word	**Gloss**
Labial		
p	*kupanga*	*to rent*

Alveolar

	Example word	**Gloss**

Palatal

Velar

	Example word	**Gloss**

Glottal

Double articulated stops

(see Part l, chapter 2, Double articulated stops)

Labialized Consonants

(see Part 1, chapter 2, Labialization and palatalization)

Geminate consonants

(see Part 1, chapter 2, Geminate consonants)

Palatalized Consonants

(see Part 1, chapter 2, Labialization and palatalization)

Syllable structures

Root initial syllable structures which occur within words

(Prefixes are irrelevant. Ignore them.)

Non-Word Final Syllables

Structure	Word	Gloss
Example: CVV	*baata*	*duck*
V		
VV		
CV		
NCV		
CVV		
NCVV		
CyVV		
NCyVV		
CwVV		
NCwVV		

Word Final Syllables

Structure	Example word	Gloss
CV		
NCV		
CyV		
NCyV		
CwV		
NCwV		

Tone

(see Part 1, chapter 2, Principles of Tone Marking in an Orthography)

Tone: Verbs

What is the tone of the infinitive prefixes:

on CVC or longer verbs?   Augment_____ Prefix _____

on -CV verb stems?   Augment_____ Prefix _____

on -VCVC verb stems?   Augment_____ Prefix _____

Describe tonal melodies on C-initial verbs, including basic patterns and any conditioned variation(s) of those patterns due to long vowels or other phonological phenomena. Give examples, with pitch traces of each pattern (use a separate page if necessary).

_____

_____

_____

_____

Note whether H is found

a) on first syllable of verb stem_____

b) on a specific syllable position (i.e., penultimate or antepenultimate)_____

c) other

In the same context as the last question, note whether H "spreads" onto following syllables._____

If so, how far does it spread?

a) One syllable?_____

b) One mora?_____

c) Two syllables?_____

d) Up through a certain syllable position (e.g., antepenultimate or penultimate)?_____

Tone: V-initial verbs

Note whether there are any different patterns, or if the normal patterns found on the V-initial verbs are merely "shifted leftward" one syllable.

_____

_____

_____

_____

Tone: Nouns

What is the tone of prefixes on CVCV (or longer) nouns?_____

Give examples, with pitch traces, of each:

_____

_____

_____

How many tone patterns are found in nouns?_____ Give examples, with pitch traces, of each.

_____

_____

_____

_____

Are there any different patterns found only in a certain noun class? If so, describe them below.

_____

_____

_____

_____

Noun Classes

Noun class prefixes with Consonant-Initial Roots

This section lists only those noun classes that exist in the language. Please record prefixes before a **consonant-initial root**.

	Kiambishi Prefix	Mfano Example	Fasiriwa Gloss
1	*m*	*mtoto*	*child*
		Mswahili	*Swahili person*
2	*wa*	*watoto*	*children*
		Waswahili	*Swahili people*
1			
1a			

	Kiambishi Prefix	Mfano Example	Fasiriwa Gloss
2			
2a			
3			
4			
5			
6			
7			
8			

	Kiambishi Prefix	Mfano Example	Fasiriwa Gloss
9			
10			
11			
12			
13			
14			
15			

	Kiambishi Prefix	Mfano Example	Fasiriwa Gloss
16			
17			
18			
19			
20			

Non-standard Noun Class Pairings

Class	Singular			Plural (non-standard)		
	Prefix	Example	Gloss	Prefix	Example	Gloss
1a/2a	-------	*bheesina*	*namesake*			
1a/8	-------	*nyabhutama*	*type of snake*	*bhi*	*bhinyabhutama*	*type of snakes*
1a/6a	------	*wasongo*	*chameleon*	*ga*	*gawasongo*	*chameleons*

Noun Class Prefixes with Vowel Initial Roots

Write the vowel list for your language in the "vowel" column. Now go back to the "prefix" column and fill in the prefixes for noun class 1. Chart the changes which take place when each vowel meets that noun class prefix.

N-Class 1 (SAMPLE from KiTharaka)

prefix	+	vowel	becomes	orthographic result	example word	gloss
mû	+	*i*	▶	*mwi*	*mwiki*	*bride*
mû	+	*î*	▶	*mwî*	*mwîtîgua*	*trustworthy person*
mû	+	*e*	▶	*mwe*	*mwekûrû*	*woman, lady*
mû	+	*a*	▶	*mwa*	*mwarimo*	*teacher*
mû	+	*u*	▶	*muu*	-----------	-----------
mû	+	*û*	▶	*mûû*	*mûûgî*	*clever person*
mû	+	*o*	▶	*mwo*	*mwokozi*	*saviour*

N-Class 1

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 2

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 3

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 4

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 5

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 6

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 7

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 8

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 9/10   (See page 177)

N-Class 11

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 12

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 13

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 14

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 15

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 16

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 17

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 18

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 19

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-Class 20

prefix	+	vowel	becomes	orthographic result	example word	gloss
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

N-class 9/10

Notes regarding class 9/10 prefixes: these may or may not be identical to each other, depending on the language. Nasals in these prefixes may be dropped out either preceding a certain phonological class of segments (in which case the filling of the noun class chart in the manual will make the environments clear) OR just in random places not definable in any phonological terms, in which case the language has a class 9a/10a which will probably include a good number of the loanwords found in class 9/10. Also, a class 9 noun which begins with Augment (**i**- or **e**-) **ny-** may either have a vowel initial root (since the **ny**- is the usual form of the 9/10 prefix preceding a vowel initial root) or actually have an <ny> initial root, since the nasal of the prefix also drops out preceding a root which begins with a nasal consonant (<m>, <n>, <ny>, or <ng'>). In order to tell whether the <ny> is actually part of the root, elicit the diminutive form ("Ask what a very small ...... would be called) and if the <ny> still shows up with the diminutive prefix, then the root should be interpreted as <ny> initial.

Prefix N	+	Phoneme	becomes	orthographic result	example word	gloss
		Voiced stops/ affricates				
	+		▶			
	+		▶			
	+		▶			
	+		▶			
		Voiced fricatives				
	+		▶			
	+		▶			
	+		▶			
	+		▶			
		Oral sonorants	▶			
	+		▶			
	+					
	+		▶			
	+		▶			

Prefix N	+	Phoneme	becomes	orthographic result	example word	gloss
		Vl. Stops/ affricates	▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			

Vl. fricatives

	+		▶			
	+		▶			
	+		▶			
	+		▶			

Nasals

	+		▶			
	+		▶			
	+		▶			
	+		▶			

Prefix N	+	Phoneme	becomes	orthographic result	example word	gloss
		Vowels				
	+	i	▶			
	+		▶			
	+		▶			
	+		▶			
	+		▶			
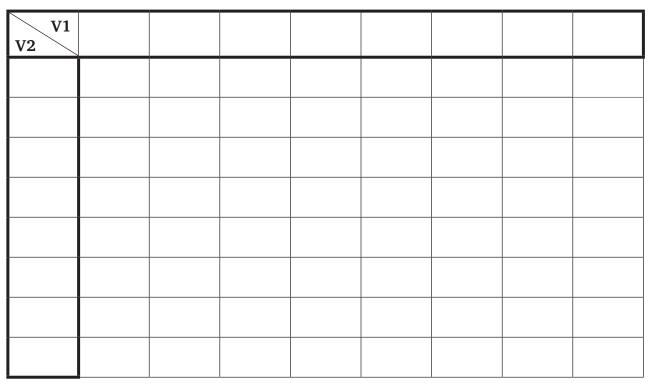
Morphophonology

Vowel Coalescence (word-internally) at morpheme boundaries

Vowel Coalescence chart (see Part 1, chapter 2, Vowel Coalescence):

This chart shows the changes that take place when two vowels occur next to each other across morpheme boundaries. List all the vowels of the language across the top and side.

The vowels listed on the left side are those which come first; those along the top come second.

V1 / V2							

Vowel Elision

(see Part 1, chapter 2, Vowel Elision (Vowels)

List examples of adjoining words which will be written as one, with an apostrophe to represent a vowel which will not be written.

_____

_____

_____

Noun Phrase

Demonstratives

Example phrases for this/these (huyu/hawa) and that/those (near) (BASIC)

	jina	this/these	that/those (near)	That/those (far)	(Other, i.e., near to addressee)
1	*mtu*	*huyu*	*huyo*	*yule*	
2	*watu*	*hawa*	*hao*	*wale*	
1					
2					
3					
4					
5					
6					
7					
8					

	jina	this/these	that/those (near)	That/those (far)	(Other, i.e., near to addressee)
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Example phrases for that/those (far) and that/those (near addressee)?

Class 1: mtu yule	*mtu -----*
Class 2: watu wale	*watu---*
Class 1	
Class 2	
Class 3	
Class 4	

Class 5	
Class 6	
Class 7	
Class 8	
Class 9	
Class 10	
Class 11	
Class 12	
Class 13	
Class 14	
Class 15	
Class 16	
Class 17	
Class 18	
Class 19	
Class 20	

Demonstratives: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Demonstrative Forms: Other

Record any other demonstratives found in the language for every noun class. There will probably be more emphatic forms than will fit on this page. Just record a couple to show the basic morphology involved:

Noun class	Any other basic demonstrative not already listed	Emphatic demonstratives - these very ones, etc.	those very ones, etc.	those very ones already mentioned, etc.
1/1a				
2/2a				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

Notes:

Demonstrative Forms, Other: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Interrogative Forms

List here the question words corresponding to the following glosses:

Noun class	Which? (ipi?)	What kind? (gani?)	How many? (ngapi?)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			

Is there a Question marker that is used before an interrogative clause, like *je* in Swahili? If so, list and give an example with a gloss:

_____

_____

Interrogative Forms: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Pronominal Forms 1

Record pronouns of the following type for every noun class:

Person or class	Prefix	Independent form: mimi, wewe, etc.	and him, etc. (nami, nawe, naye, etc.)	emphatic copular pronoun (he is the one who) (ndimi, ndiwe, ndiye, etc.)	negative copular pronoun (hapana, hakuna, etc.)
1s					---------
2s					---------
3s (1/1a)					
1p					----------
2p					----------
3p (2/2a)					
NCl.3					
4					
5					
6					
7					
8					
9/9a					
10/10a					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

Pronominal Forms 1

Pronominal Forms 1: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Pronominal Forms 2: Quantificational

Record pronouns of the following type for every noun class:

Person or class	Prefix	All/whole (of it/them)/ e.g., sote, nyote, yote, pote, etc.	Both of them (note if any other lexicalized forms e.g., all three of them, etc.)	Any (of them) yeyote, etc.
1s	------	-------	----	--------
2s	-------	-------	----	--------
3s (1/1a)				
1p				---------
2p				---------
3p (2/2a)				
NCl.3				
4				
5				
6				
7				
8				
9/9a				
10/10a				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

Pronominal Forms 2: Quantificational: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Pronominal Forms 3: Set pronouns and related forms

Record pronouns of the following type for every noun class:

Person or class	Prefix	Another of their same set - mwenzao, wenzao, etc.	"Self" alone	Other "mwingine"
1s		--------		--------
2s		--------		--------
3s (1/1a)				
1p		---------		---------
2p		---------		---------
3p (2/2a)				
NCl.3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Pronominal Forms 3, Set Pronouns and related forms: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Adjectives (Qualifiers)

	jina	kivumishi (konsonanti)	kivumishi (irabu)	kivumishi kingine	kivumishi kingine	Prefix List
1	*mtu*	*mrefu*	*mwerevu*			*m-, mw-*
2	*watu*	*warefu*	*werevu*			*wa-, w-*
1						
1a						
2						
2a						
3						
4						
5						
6						
7						
8						
9						
9a						
10						
10a						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Adjectives, Qualifiers: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Adjectives (Quantifiers)

	jina	wote	yeyote wowote			Prefix list
1	*mtu*	*-----*	*yeyote*			*ye-*
2	*watu*	*wote*	*wowote*			*wo-*
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Adjectives, Quantifiers: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Associative Constructions

	Noun	of -a	child mtoto	
1	*mwalimu*	*wa*	*mtoto*	
2	*walimu*	*wa*	*mtoto*	
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Associative Constructions: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Diminutives and Augmentatives

N-Class 1 (Sample from Kitharaka)

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV	*mûkûrû*	*older man*	*kamûkûrû*	*kîmûkûrû*
VCV	*(mw)ana*	*child*	*kamwana*	*kîmwana/kîana*
CV	*(mu)ntû*	*person*	*kamuntû-*	*kîmuntû*

N-Class 1

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 2

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 3

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 4

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 5

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 6

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 7

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 8

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 9

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 10

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 11

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 12

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 13

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 14

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 15

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 16

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 17

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 18

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 19

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

N-Class 20

Syllable Pattern On Root	Sample Noun	Gloss	Diminutive	Augmentative
CVCV				
VCV				
CV				

Diminutives and Augmentatives: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Numerals

	jina	1	2	3		Prefix List
1	*mtu*	*mmoja*	*--*	*--*	*--*	*m-*
2	*watu*	*--*	*wawili*	*watatu*		*wa-*
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Numerals: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Locatives

(Sample from Niha)

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV	*nyumba*	*pa=nyumba*	*ku=nyumba*	*mu=nyumba*
VCV	*ikwi*	*pikwi*	*kwikwi*	*mwikwi*
CV	*mbwa*	*pa=mbwa*	*ku=mbwa*	*mu=mbwa*

N-Class 1

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 2

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 3

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 4

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 5

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 6

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 7

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 8

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 9

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 10

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 11

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 12

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 13

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 14

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 15

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 19

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

N-Class 20

Syllable Pattern On Root	Sample Noun	Loc 16	Loc 17	Loc 18
CVCV				
VCV				
CV				

Locatives: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Locatives: Spelling Rules

(Please give a few examples)

Possessive pronouns

	Noun	my wangu	your wako	his wake	our wetu	your (pl) wenu	their wao
1	*mtu*						
2	*watu*						
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Possessive Pronouns: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Possessive Pronouns: Spelling Rules

(Please give a few examples)

Use of the Augment on Nouns

Matumizi Ya Viambishi Vya Viambishi Vya Nomino

1. Does the augment (i.e., pre-prefix, the initial vowel of the nominal prefix) usually appear in nouns used in isolation (i.e., nouns listed without any context)?   YES/NO (Circle one.)

   Give two or three examples (one with each different augment vowel) with glosses:

   _____

   _____

   _____

   Note any general exceptions, e.g., a certain class in which the augment is used differently (you may cross reference this with another chart if desired):

   _____

   _____

2. If the augment is generally used on nouns in isolation, can it be removed? If so, can people describe any difference in meaning between the form with and without the augment?

   _____

   _____

   _____

3. Is the augment used in the noun phrase on words other than nouns, e.g., adjectives, numerals, etc.? To determine this, translate the following:

Mti mrefu: _____

Ni mti: _____

Ni mti mrefu.: _____

Mti ni mrefu.: _____

Mti mrefu mmoja ulianguka.: _____

Waliangusha miti yote.: _____

Hawakuangusha miti yoyote.: _____

Verbs (Verbal Affixation)

Inflectional

- subject prefixes (for participants and for classes)
- object prefixes (for participants and for classes)
- relative prefixes (for participants and for classes)
- T/A markers
- negative markers
- any other inflectional elements particular to a language

Derivational

- Verb-verb:
- Productive derivational suffixes: causative, passive, applicative, reciprocal
- Combinations of these
- Changes in verbal aspectual suffixes because of the use of (combinations of) derivational suffixes
- Verb-Noun: agentive nouns derived from verbs

Verb-Adjective: adjectives derived from verbs

Copular Forms

Noun Class	Noun w/gloss (e.g., kitabu - book)	It is a….. (e.g., Ni kitabu.)	It is not a……. (e.g., Si kitabu.)	This is a …., not a ….) (e.g., Ni kitabu, siyo daftari.)
1/1a				
2/2a				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

Notes:

Copular Forms: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Copular Forms: Spelling Rules

(Please give a few examples)

Morphophonological Processes In Verb Stems With Perfective Final Suffix

In the first part of the chart, for each consonant in the phonemic inventory of your language, find a verb stem ending with that consonant. List the consonant in the first column, the infinitive and its gloss in the second and third columns, and a verb form with perfective ending (usually can be elicited using *me-* tense in Swahili, e.g., *ameketi*, etc.) in the fourth column. Note in the last column any changes that occur in the stem-final consonant. For example, in Kifuliiru, in a verb stem ending with <k>, the <k> is changed to <s> in the perfective form: *kubiika* 'to put', *abiisiri* 'he has put'.

For the second part of the chart, find a verb ending with each of the extensions listed and note any changes in the extension in the perfective forms. In some extensions, the final consonant may drop before the "perfective" ending. Vowel coalescence may then take place.

Final Consonant	Infinitive	Gloss	Form with perfective ending (use 1 sg. subject)	Change in final Consonant e.g., None, k→ s, g →z, etc.

Final Consonant	Infinitive	Gloss	Form with perfective ending (use 1 sg. subject)	Change in final Consonant e.g., None, k→ s, g →z, etc.
Extension				
Appl. (-ir/-er)				
Appl. (-ir/-er)				
Recip. (-ana)				
Causative (short)				
Causative (long)				
Separative (-ol/-ul)				
Separative (-ol/-ul)				

Notes:

Verb Stems with Perfective Final Suffix: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Imperative Plural

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive *kushiba* has a two-syllable stem, while *kula* has a one-syllable stem, and *kukalia* has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive-Tone cl. (H/L, H w/ VV, etc.)	Gloss	Imperative PL	Imperative PL w/H tone obj	Imperative PL w/L tone obj
3-syll. stem					
3-syll. stem					
3-syll. stem					
Neg.					

2-syll. stem					
2-syll. stem					
2-syll. stem					
Neg.					

1-syll. stem					
1-syll. stem					
1-syll. stem					
Neg.					

Verbal Tense/Aspect, Imperative Plural: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Imperative Singular

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive *kushiba* has a two-syllable stem, while *kula* has a one-syllable stem, and *kukalia* has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive-Tone cl. (H/L, H w/VV, etc.)	Gloss	Imperative	Imperative w/H tone obj	Imperative w/L tone obj
3-syll. stem					
3-syll. stem					
3-syll. stem					
Neg.					

2-syll. stem					
2-syll. stem					
2-syll. stem					
Neg.					

1-syll. stem					
1-syll. stem					
1-syll. stem					
Neg.					

Imperative Singular: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Imperative Singular: Spelling Rules

(Please give a few examples)

Verbal Tense/Aspect: Simple Subjunctives

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive *kushiba* has a two-syllable stem, while *kula* has a one-syllable stem, and *kukalia* has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive-Tone cl. (H/L, H w/VV, etc.)	Gloss	Imperative	Imperative w/H tone obj	Imperative w/L tone obj
3-syll. stem					
3-syll. stem					
3-syll. stem					
Neg.					
2-syll. stem					
2-syll. stem					
2-syll. stem					
Neg.					
1-syll. stem					
1-syll. stem					
1-syll. stem					
Neg.					

Simple Subjunctives: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Subjunctives – Distal

(e.g., go and do x - uende ukafanya....) Mark the placement of any H, rising, or falling tones on all forms. "Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive *kushiba* has a two-syllable stem, while *kula* has a one-syllable stem, and *kukalia* has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive-Tone cl. (H/L, H w/VV, etc.)	Gloss	Imperative	Imperative w/H tone obj	Imperative w/L tone obj
3-syll. stem					
3-syll. stem					
3-syll. stem					
Neg.					

2-syll. stem					
2-syll. stem					
2-syll. stem					
Neg.					

1-syll. stem					
1-syll. stem					
1-syll. stem					
Neg.					

Subjunctives, Distal: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Prefixes: Subject

Using your list of example nouns from each noun class, elicit a simple subjunctive verb form with a subject marker that agrees with each class. It may be necessary to switch verbs occasionally for reasons of semantic non-compatibility with subjects of certain classes, but make an effort to keep the verb in each column the same if possible (i.e., ideally, use only three different verbs for this chart - one verb for each column). A passive or "neuter" verb form, if a suitable one can be found, may be more likely to be semantically compatible with multiple subjects, e.g., "may x be found," or "may x appear," etc. Using *a simple subjunctive form* ensures that the subject prefix appears right next to the verb root, e.g., 1st singular forms in Swahili might be **nifike** *that/may I arrive,* **niende** *that/may I go,* **niombe** *that/may I pray*, etc.

Person or class	Subj Prefix (basic)	C-initial verb root (e.g., -fika)	V-initial verb root (e.g., -enda)	V-initial verb root (e.g., omba)	V-initial verb	V-initial verb	V-initial verb
Inf/ gloss							
1s							
2s							
3s							
1p							
2p							
3p							
NCl.3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							

Person or class	Subj Prefix (basic)	C-initial verb root (e.g., -fika)	V-initial verb root (e.g., -enda)	V-initial verb root (e.g., omba)	V-initial verb	V-initial verb	V-initial verb
16							
17							
18							
19							
20							
21							
22							
23							

If there is any vowel coalescence going on between a subject prefix and a vowel-inital root, find a verb to represent each vowel which can be found initially in a verb root, and fill the entire paradigm (i.e., one verb which begins with <i>, another verb which begins with <e>, another beginning with <o>, etc.). Use another copy of this chart if necessary.

If any variation in the form of the prefixes is conditioned by the following (verb root-initial) consonant, make note of that below, pointing out an example (above) for each variation. Such conditioned variants will normally only involve a prefix that consists only of a nasal consonant (usually first person singular, e.g., first person singular prefix **n**- appears as **m**- when it occurs directly preceding a verb root beginning with a bilabial consonant.

Notes:

_____

_____

_____

Verbal Prefixes, Subject: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Prefixes, Subject: Spelling Rules

(Please give a few examples)

Verbal Prefixes: Object Prefixes

Fill this chart using transitive verbs with first plural subject so that each gloss is "*we ... [ourselves]*," "*we .... [you (sg)]*," "*we .... [him]*," etc. The exception to this pattern is that in the elicitation of first person object markers, a second person plural subject should be used ("*you ...me*," "*you......us*"). The RFLX row should be done with first person plural subject—"*we .... ourselves.*" For each vertical column, use only one verb root, and do not change tenses. The form should be either a subjunctive or a simple past, corresponding to the Kiswahili *tulijipiga, tulikupiga, tulimpiga*, etc.

Fill the entire paradigm for a verb representing each initial vowel you find in verb roots (i.e., one verb which begins with *i*, another verb which begins with *e*, etc.). If any variation in the form of the prefixes is conditioned by the following (verb root-initial) consonant, make note of that below, giving an example for each variation. Such conditioned variants will normally only involve a prefix that consists only of a nasal consonant (usually first person singular).

Note any object prefixes that seem to have H tone. An H tone with a first person singular object of the shape N- will often be realized on a preceding or syllable.

Person or class	Obj Prefix	C-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root
Inf/ gloss							
RFLX							
1s							
2s							
3s							
1p							
2p							
3p							
NCl.3							
4							
5							
6							
7							
8							
9							
10							
11							

Person or class	Obj Prefix	C-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root	V-initial verb root
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

If it is possible to have more than one object prefix in the verb, give some examples: we gave it (food) to him, we gave it (food) to it (e.g., dog, cow, etc.). If this is not possible, see if it is possible to have two objects if one of them is reflexive, if one of them is first person (singular or plural), if one of them is second person (singular or plural).

Notes:

_____

_____

Verbal Prefixes, Object: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Prefixes: Object Relative

Using your list of example nouns from each noun class, elicit a simple past verb form with a first person plural subject (*we*) with a relativized object that agrees with each class. These will be equivalent to phrases like **ambaye tulimwona/ tuliyemwona** *(the person) whom I saw*, **ambao tuliwaona/tuliowaona** *(the people) whom we saw*, **tuliyoiona** (cl.) *which we saw*, ... etc. In a language other than Swahili, the relativized object may or may not need to be repeated just before the verb stem. In order to elicit these, it may be helpful to put the clause into a sentence such as: This is the ....(insert noun of the appropriate class) which we saw. List below only the part of the sentence translatable as "*whom/which we saw.*"

Noun class	whom/which we saw
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
19	
20	

Notes:

_____

_____

Verbal Prefixes, Object Relative: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Prefixes: Subject Relative

Using your list of example nouns from each noun class, elicit a simple past verb form with a relativized subject agreeing with each person and class. These will be equivalent to phrases like **niliyeanguka** *I who fell,* **uliyeanguka** *you who fell,* **aliyeanguka** *he who fell, ...* **ulioanguka** *(the tree) which fell,* etc. List these in the first column. In the second column, list the same forms but in simple future tense **nitakayeanguka** *I who will fall,* etc.. In the third column, list negative forms: **nisiyeanguka**, **usiyeanguka**, etc.

Person or class	Rel Prefix	Past Relative	Future Relative	Negative Relative
Inf/gloss				
1s				
2s				
3s				
1p				
2p				
3p				
NCl.3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Notes:

Verbal Prefixes, Subject Relative: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Future 1

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Fut 1 Pfx	Fut 1	Fut 1 w/H tone obj (cl. ___)	Fut 1 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg						

Notes:

Verbal Tense/Aspect, Future 1: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Future 2

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Fut 2 Pfx/	Future 2	Future 2 w/H tone obj (cl. ___)	Future 2 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

Verbal Tense/Aspect, Future 2: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Future 3

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Fut 3 Pfx/	Future 3	Future 3 w/H tone obj (cl. ___)	Future 3 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Future 3: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect, Future 3: Spelling Rules

(Please give a few examples)

Verbal Tense/Aspect: Past 1

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Past 1 Pfx	Past 1	Past 1 w/H tone obj (cl. ___)	Past 1 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Past 1: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Past 2

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Past 2 Pfx	Past 2	Past 2 w/H tone obj (cl. ___)	Past 2 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

Verbal Tense/Aspect, Past 2: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Past 3

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Past 3 Pfx	Past 3	Past 3 w/H tone obj (cl. ___)	Past 3 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

Verbal Tense/Aspect, Past 3: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Past 4

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Past 4 Pfx	Past 4	Past 4 w/H tone obj (cl. ___)	Past 4 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

Verbal Tense/Aspect, Past 4: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Present - Unmarked

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Present Pfx	Present - unmarked	Present - unmarked w/H tone obj (cl. ___)	Present - unmarked w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

Verbal Tense/Aspect, Present Unmarked: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Present 2

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Pres 2 Pfx	Present 2	Present 2 w/H tone obj (cl. ___)	Present 2 w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Present 2: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Other: _____ (Describe here)

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below.

	Infinitive	Gloss	Pfx Aux	_____	_____ w/H tone obj (cl. ___)	_____ w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						

2-syll.						
2-syll.						
2-syll.						

1-syll.						
1-syll.						
1-syll.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Other: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Anterior

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Anterior Pfx/Sfx	Anterior	Anterior w/H tone obj (cl. ___)	Anterior w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Anterior: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Past Anterior

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Past Anterior Pfx/Sfx	Past Anterior	Past Anterior w/H tone obj (cl. ___)	Past Anterior w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Past Anterior: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Present Progressive

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Progressive Pfx/Aux	Progressive	Progressive w/H tone obj (cl. ___)	Progressive w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Present Progressive: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Present Persistive

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" refers to all parts of the verb to the right of the infinitive prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Pres Pfx	Present persistive	Present persistive w/H tone obj (cl. ___)	Present persistive w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Present Persistive: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Verbal Tense/Aspect: Consecutive

Mark the placement of any H, rising, or falling tones on all forms.

"Stem" below refers to all parts of the verb to the right of any object prefix. E.g., The infinitive **kushiba** has a two-syllable stem, while **kula** has a one-syllable stem, and **kukalia** has a three-syllable stem. If there are two (or more) tone classes of verbs in your language, include one of each. If a verb with a long vowel exhibits a slightly different surface tonal pattern, include such a verb in the table below, along with the negative form of one verb of each syllable count.

	Infinitive	Gloss	Consecutive Pfx	Consecutive	Consecutive w/H tone obj (cl. ___)	Consecutive w/L tone obj (cl.___)
3-syll.						
3-syll.						
3-syll.						
Neg.						

2-syll.						
2-syll.						
2-syll.						
Neg.						

1-syll.						
1-syll.						
1-syll.						
Neg.						

Notes:

_____

_____

_____

Verbal Tense/Aspect, Consecutive: Spelling Rules

(Please give a few examples)

Decisions and rationale:

Residue:

Vowel Harmony in Extensions

Choose simple -CVC- verbs for the chart below. The V in each verb should represent each different phonemic vowel in your language. Choose a verb of each vowel to represent each verbal extension (derivational suffix) listed below. If possible, use a single verb for each column. However, various verbs may need to be used in a single column in order to find a root that is semantically compatible with the suffix.

Watch for any vowel harmony things that may be happening. The extension may harmonize with the stem, or sometimes in the causative, the stem harmonizes with the extension.

The examples below are not all naturally occurring forms in Swahili, but are intended only to model the suffix.

Stem Vowel →	a	gloss	e	gloss
Example ↓				
tenda				
tendea				
tendesha				
tendekeza				
tendeka				
tendana				
tendua				
tenduka				

Stem Vowel →	i	gloss	o	gloss
Example ↓				
tenda				
tendea				
tendesha				
tendekeza				
tendeka				
tendana				
tendua				
tenduka				

Stem Vowel →	u	gloss		gloss
Example ↓				
tenda				
tendea				
tendesha				
tendekeza				
tendeka				
tendana				
tendua				
tenduka				

List appropriate vowels in the following fill-in-the-blank statements. The first blank in each statement may have two or more vowels listed, but only one vowel should be listed in the final blank of each statement. E.g., the first statement may look like this:

In a verb with a stem vowel of __i, u, or a_____ the vowel of a front vowel suffix is realized as____i_____.

(A language with five vowels will need to fill only the first two in each group. Languages with seven vowels will probably have three different forms of each type of suffix.)

Front vowel suffixes (-ir/er, -ik/-ek, etc.)

In a verb with a stem vowel of _____ the vowel of a front vowel suffix is realized as_____.

In a verb with a stem vowel of _____ the vowel of a front vowel suffix is realized as_____.

In a verb with a stem vowel of _____ the vowel of a front vowel suffix is realized as_____.

Back vowel suffixes (-uk/-ok, -ul/-ol, etc.)

In a verb with a stem vowel of _____ the vowel of a back vowel suffix is realized as_____.

In a verb with a stem vowel of _____ the vowel of a back vowel suffix is realized as_____.

In a verb with a stem vowel of _____ the vowel of a back vowel suffix is realized as_____.

Chart Checklist

Page	Chart	Completed	Consultant Checked
Vowels			
5	Table 1: Vowel Inventory	_____	_____
5	Table 2: V1 - V2 combinations chart	_____	_____
6	Table 3: Long Vowel occurrence in disyllabic word roots	_____	_____
7	Table 4: Vowel Length	_____	_____
8	Table 5: Vowel Harmony	_____	_____
Consonants			
9	Consonants: Root-initial	_____	_____
11	Consonants: Root-medial	_____	_____
12	List of examples for each consonant phoneme	_____	_____
14	Double articulated stops	_____	_____
14	Labialized Consonants	_____	_____
15	Geminate consonants	_____	_____
15	Palatalized Consonants	_____	_____
Syllable structures			
16	Non-Word Final Syllables	_____	_____
17	Word final syllables	_____	_____
Tone			
18	Tone: Verbs	_____	_____
19	Tone: V-initial verbs	_____	_____
19	Tone: Nouns	_____	_____
Noun Classes			
20	Noun class prefixes with Consonant Initial Roots	_____	_____
24	Non-standard Noun Class Pairings	_____	_____
25	Noun Class Prefixes with Vowel Initial Roots	_____	_____
Morphophonology			
39	Vowel Coalescence (word-internally) at morpheme boundaries	_____	_____
39	Vowel Elision	_____	_____

Page	Chart	Completed	Consultant Checked
Noun Phrase		_____	_____
40	Example phrases for this/these (huyu/hawa) and that/those (near) (BASIC)	_____	_____
42	Example phrases for that/those (far) and that/those (near addressee)?	_____	_____
45	Demonstrative forms: other	_____	_____
47	Interrogative Forms	_____	_____
49	Pronominal Forms 1	_____	_____
		_____	_____
		_____	_____
		_____	_____
		_____	_____
		_____	_____

	Complete	Consultant Checked
Interrogative		
Pronominal forms -1		
Pronominal forms -2		
Pronominal forms -3		
Adjective		
Associative Constructions		
Diminutives and Augmentatives		
Enumeratives		
Locatives		
Possessive pronouns		
Verbs:		
Copular forms		
Copular forms part 2		
Morphophonological processes…Perfective		
Verbal Mood:Imperative plural		
Verbal Mood:Imperative singular		
Verbal Mood:Subjunctive		
Verbal Mood:Subjunctive 2		
Verbal Prefixes: Subject.		
Verbal Prefixes: Subject 2		
Verbal Prefixes: Object		
Verbal Prefixes: Object Relat.		
Verbal Prefixes: Subject Relat.		
Verbal Tense: Future 1		
Verbal Tense: Future 2		
Verbal Tense: Future 3		
Verbal Tense: Future 4		
Verbal Tense: Past 1		
Verbal Tense: Past 2		
Verbal Tense: Past 3		
Verbal Tense: Present or unmarked		
Verbal Tense: Present 2		
Verbal Tense-Aspect/other		
Verbal Tense-Aspect Anterior		
Verbal Tense-Aspect Past Anterior		
Verbal Tense-Aspect Progressive		
Verbal Tense-Aspect Present Persistive		
Verbal Tense-Consecutive		
Vowel Harmony in Extensions		
Conditional (if it exists)		
Regret?		

References

Anderson, Stephen C. and Christian Alomofor. 2005. Awing Orthography Guide. Cameroon: SIL.

Bastin, Yvonne. 2003. The interlacustrine zone. In Derek Nurse and Gérard Philippson (eds.), 501–527.

Bird, Steven. 1998. *Strategies for representing tone in African writing systems: A critical review.* Edinburgh: University of Edinburgh.

Bird, Steven. 2000. Orthography and identity in Cameroon. *Notes on Literacy* 6(1–2):3–34.

Botne, Robert. 2003. Lega. In Derek Nurse and Gérard Philippson (eds.), 422–445.

Bresnan, Joan and Samuel A. Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13:181–254.

Burquest, Donald A. 1993. *Phonological analysis: A functional approach.* Dallas: SIL.

Edmondson, Jerold A. and Donald Burquest. 1998. *A survey of linguistic theories.* Dallas: SIL.

Eira, Christina. 1998. Authority and discourse: Towards a model for orthography selection. *Written Language and Literacy* 1(2):171–224.

Gardner, William L. 2001. Consonant mutation in Shona languages. In Thomas Crowell and Janet Harthan (eds.), *Working papers 2*, 51–59. Nampula, Mozambique: Sociedade Internacional de Linguística.

Gardner, William L. 2005. *Overview of Bantu languages.* Tutorial paper, Fuller Theological Seminary.

Gowlett, Derek. Zone S. In Derek Nurse and Gérard Philippson (eds.), 609–638.

Gudschinsky, Sarah. 1973. *A manual of literacy for preliterate peoples.* Dallas: SIL.

Hyman, Larry M. 2003. Segmental phonology. In Derek Nurse and Gérard Philippson (eds.), 42–58.

Katamba, Francis. 2003. Bantu nominal morphology. In Derek Nurse and Gérard Philippson (eds.), 103–120. .

Kithinji, Stephen, Nick Lunn, and Leila Schroeder. 1999. *Tharaka orthography.* Manuscript.

Kroger, Paul. 1990. Lexical phonology and the rebirth of the phoneme. *Notes on Linguistics* 50:11–24.

Kutsch-Lojenga, Constance. 1993. The writing and reading of tone in Bantu languages. *Notes on Literacy* 19(1):1–19.

Kutsch-Lojenga, Constance. 2003. Bila. In Derek Nurse and Gérard Philippson (eds.), 450–473.

Leitch, Myles. 2000. *SHOEBOX Babole dictionary: Computerized Babole-French research dictionary*. Libreville: SIL.

Leitch, Myles. 2003. Le systeme babole. In F. Idiata (ed.), *Les classes nominales et leur semantisme dan les langues bantoues du nord-ouest*, Munich: Lincom Europa.

Maddieson, Ian. 2003. The sounds of the Bantu languages. In Derek Nurse and Gérard Philippson (eds.), 15–41.

Mattingly, I. G. 1992. Linguistic awareness and orthographic form. In R. Frost and L. Katz (eds.), *Orthography, phonology, morphology, and meaning,* 11–26. Amsterdam: North-Holland.

Moe, Ronald. 2003. *Developing an orthography for a Bantu language.* Manuscript.

Moe, Ronald and Samuel Mubbala. 2006. *Lugwere writers' guide.* Entebbe: SIL.

Nurse, Derek and Gérard Philippson. 2003. *The Bantu languages.* London: Routledge.

Nzogi, Richard. 2005. *Lugwere orthography statement.* Manuscript.

Odden, David. 2003. Rufiji-Ruvuma (N10, P10-20). In Derek Nurse and Gérard Philippson (eds.), 143–163.

Reeder, JeDene. 1998. *Vowel harmony in Pagibete verbs.* Manuscript.

Roberts, David. 2008. *Thirty Years of Tone Orthography Testing, Journal of West African Languages 200 XXXV.1-2.*

Schadeberg, T. C. 1999. Katupha's law in Makhuwa. In J. M. Hombert and L. M. Hyman (eds.), *Bantu historical linguistics: theoretical and empirical perspectives*, 379–394. Stanford: CSLI Publications.CSLL

Schadeberg, T. C. 2003. Historical linguistics. In Derek Nurse and Gérard Philippson (eds.), 143–163.

SIL International. 2002. Developing an orthography. *In LinguaLinks version 5*. CD-ROM. Dallas: SIL.

SIL International. 2002. How to make spelling choices. *In LinguaLinks version 5*. CD-ROM. Dallas: SIL.

Smalley, William A. (ed.) 1964. *Orthography studies*. London: UBS.

Snider, Keith L. 1992. Grammatical tone and orthography. *Notes on Literacy* 18(4):25–30.

Snider, Keith L. 1998. Orthography design and development. Paper presented at the SIL

Literacy Consulting Skills Seminar, Yaounde, Cameroun. August 29–29, 1998.

Snider, Keith L. 2001. Linguistic factors in orthography design. In Ngessimo M. Mutaka and Sammy B. Chumbow (eds.), *Research material in African linguistics: Focus on Cameroon,* 323–332. Koln: Rudiger Köppe Verlag.

Snider, Keith L. 2005. You can have your cake and eat it, too: An orthography that meets the needs of both mature and immature readers. Lecture notes for an orthography class at the University of North Dakota, June 29, 2005.

Ssemakula, Mukasa E. 2005. *A Primer on Speaking and Writing Luganda,* http://www. buganda.com/Luganda.htm#elem) (accessed April 27, 2005).

Stegen, Oliver. 2005. Vowel length in eastern Bantu orthographies. Paper presented at the SIL Bantu orthography meeting Dallas, November 7–12, 2005.

Van Dyken, Julia R. and Constance Kutsch-Lojenga. 1993. Word boundaries: Key factors in orthography development. In Rhonda Hartell (ed.), *Alphabets of Africa,* 3–20. Dakar: UNESCO.

Van Otterloo, Karen and Roger Van Otterloo. To appear. *Determining word boundaries.* Dallas: SIL.