

Researching and Documenting the Languages of Tanzania

Henry R.T. Muzale and Josephat M. Rugemalira
University of Dar es Salaam

This paper describes the challenges that researchers have encountered during six years of implementing a research and documentation project for the languages of Tanzania. It discusses the methods evolved by the project researchers for the production of a language atlas for Tanzania and presents preliminary results from the research. The results show that the language with the most native speakers, Sukuma, has twice as many as its closest rival, Kiswahili. The paper also presents an account of the research for documenting the grammars and vocabularies of the languages of Tanzania. The expected impact of this particular form of documentation, as well as the limits, are discussed. It is argued that a language needs to be unchained from politically imposed shackles in order for a society to reap the full benefits of its cultural resources.

1. INTRODUCTION. This paper focuses on the challenges that researchers have encountered during the past seven years of pursuing the objectives of the Languages of Tanzania Project. It describes the efforts to document the grammars and vocabularies of some of the languages of Tanzania against the background of a political agenda that has sought to promote one national language and ignore, even suppress, all the other Tanzanian languages. We discuss the various attempts to evolve an efficient method for producing a language atlas for Tanzania and the current status of the project. We also present the research results. We will show that in the absence of language data from the national population census it is still possible to obtain reliable information by other methods.

2. BACKGROUND TO THE PROJECT. The research and documentation project for the languages of Tanzania was launched in 2001 at the University of Dar es Salaam (Muzale and Rugemalira 2001, Legere 2002). The project has two major objectives:

1. To produce a language atlas for the country, indicating the languages spoken in Tanzania, the number of speakers for each language, and geographical distribution of the languages.
2. To produce a vocabulary list/dictionary and a grammar for each of the languages of Tanzania.

The significance of these objectives can be better appreciated against the somewhat peculiar sociolinguistic situation in Tanzania. Besides English and the national lingua franca, Kiswahili, there are over a hundred other native languages that are not accorded any

official status. It is worth noting that there does not exist any institution for researching or promoting any of the ethnic community languages (ECLs) of Tanzania. The University of Dar es Salaam does not have a department of African Languages and Literatures as many similar universities across the continent do. In contrast, there exist both a department of Kiswahili and the Institute for Kiswahili Research at the University. In addition, there exists a National Kiswahili Council established by an act of parliament.

The ethnic community languages exist in a hostile political environment. Several regulations and policies restrict the domains of use for the ECLs. They are not permitted in the schools,¹ in the media, or in politics. Television and radio license regulations prohibit the use of ECLs in any programming, even though the recent liberalization of mass media ownership (away from the previous state monopoly) and advances in technology have supported the rise of artists singing in a few ECLs. It is practically impossible to obtain permission to register a newspaper that uses an ECL; besides, using an ECL in political campaigns would constitute a sufficient irregularity for the court to nullify an election. This almost comprehensive ban on the ECLs helps the state to maintain a strong grip on the lives of the people and limit the space for divergent ideas. It certainly accelerates the demise of the ECLs, as their use is confined to the domestic realm, and they are eclipsed by Kiswahili.

Existing information on the number and names of the ECLs is based on the 1947, 1957, and 1967 population census results. In those years, the census included a question on ethnic identity (Egero and Henin 1971). Subsequent censuses have eliminated questions related to ethnicity and language. As a result, information regarding the number of speakers for the various languages, as well as their sociolinguistic profiles, is limited to studies of individual languages and is, at times, conjectural (cf. Mekacha 1993; Caston et al. 1996, 1997; Turner et al. 1998; Msanjila 1999). Work on the grammar, vocabulary, and oral literatures covers only a small proportion of the languages (Maho and Sands 2002). Most written materials in the ECLs are religious—the Bible or parts thereof, hymns, catechetical literature, and prayer books. Even in the religious domain the Swahili onslaught is evident: observation in one rural church in Northern Tanzania, during a three-hour Sunday service, showed that only one-third of the time was devoted to use of the local ECL (involving a few hymns and prayers). Most hymns were led by the local youth choir in Kiswahili; all scripture readings were from the Kiswahili translation, although a recent New Testament translation in the local language exists (Bible Society of Tanzania 2000a). The sermon was delivered in Kiswahili by a guest pastor (Rugemalira, field notes 2004).

¹ Even if many children come to school without any knowledge of Kiswahili and, at times, a local ECL may be (illegally) used, it is with the understanding that proper education/knowledge is conveyed through Kiswahili—which is **the** language (*lugha*), as opposed to an ECL which is a mere *kilugha*, i.e., something less than a language, a despicable dialect. One of the disciplinary offences in a primary school is to speak *kilugha*. In the secondary school, of course, **the** language is English (see below).

3. DOCUMENTATION OF THE GRAMMAR AND VOCABULARY.

3.1 VOCABULARY.

One of the most successful outcomes of the project has been the production of lexicons (see Appendix 1). The design of the form and content of the lexicons, originally conceived to be classified word lists in English and the respective ECL, has evolved to embrace a wider audience and bigger challenges (Rugemalira 2004). Instead of targeting the audience of linguists and related researchers alone, the project has sought to primarily address the needs of native speakers of the ECLs. The introduction to the Ruhaya dictionary puts it succinctly:

[The dictionary] is meant for two major groups. The first group consists of the native speakers of Ruhaya. To them, this is a reservoir for the lexemes of the language with their equivalents in Kiswahili and English. The second group includes researchers in linguistics who will find valuable data for linguistic analysis and comparative studies. (Muzale 2006: xxxiii)

Accordingly, an alphabetical word list was compiled. Initially the list contained 3000 items in English, with Kiswahili equivalents. Eventually a 5000-item word list in Kiswahili, with English equivalents, was put together with the help of several existing lists (see Appendix 2). This now constitutes the basic tool for eliciting vocabulary. When it is used while working on a particular language, the list may be expanded in various ways, such as using the SIL semantic domains list (Moe 2002), or following up on a derivational pattern like that of the verb suffixes in Bantu. In many cases the particular language application of this tool comes out with fewer items because some items in the basic list are inappropriate or nontranslatable (e.g., 'coconut' in many non-coastal languages) or because several English/Kiswahili items translate into one item in the ECL. The result is a trilingual wordlist with several possibilities for the final output. The standard output consists of a two-part wordlist of the form ECL-Kiswahili-English and English-ECL-Kiswahili. Some outputs consist of only the first part, and it may be possible to put out versions with Kiswahili as the entry language. Bilingual ECL-Kiswahili or ECL-English and their reverse may also be produced.

The basic word list was prepared in the MS Excel format, and most of the products so far published are in this format. As the research team gained experience, it became possible to start working with a different format, SIL's Toolbox, which allows for the export of data into the standard dictionary format in MS Word. Extracts from the Ciruuri Lexicon (Massamba 2005) and the Ruhaya Dictionary (Muzale 2006) are presented to exemplify the two formats.

EXTRACT 1. *Ciruuri* Lexicon (Part I) in Excel format

Pf	<i>Ciruuri</i>	WC	<i>Kiswahili</i>	English
	saangirisyâ	v	<i>ingilia kula chakula kilichokusudiwa mtu mmoja</i>	join in eating food meant for one person
	saánja	v	<i>tokota; chemka</i>	simmer; boil up
	saanjága	v	<i>ponda</i>	crush by pounding; pulver- ize
	saára	v	<i>tahiri</i>	circumcise
i	saaro	n	<i>jando</i>	period of circumcision (for male)
	saarúra	v	<i>atua</i>	split/crack
i	saasi	n	<i>nzi</i>	housefly
	saasíkana	v	<i>changanyika</i>	mingle; be mixed
	saásya	v	<i>changanya</i>	mix
i	saatu	n	<i>ngege, sato</i>	tilapia
	saáya	v	<i>hara</i>	have diarrhea
	sabháánjuka	v	<i>tika</i>	move about in a container (of liquid)
i	sabháato	n	<i>sabato</i>	Sabbath
i	sábhi	n	<i>kifaru</i>	rhinoceros
i	sabhúni	n	<i>sabuni</i>	soap

Where: *n* = noun; *pf* = prefix; *v* = verb; *wc* = word class

When published, the lexicon may turn out to be the first or only publication in the language. It can easily act as a standard, particularly with regard to orthographic conventions. It also raises the value of the language among the speakers themselves and among neighboring language communities. Furthermore, the trilingual format is in part necessary, because the metalanguage has not yet been developed for the ECLs. But it is also a strong link with other languages, which native users of ECLs find very useful.

The standardizing aspect of the word lists as regards orthography has raised a number of challenges. The first concerns the extent to which it is desirable and possible to develop common orthographic conventions for all the languages of Tanzania. Ideally, a uniform set of conventions that would cover all the languages would be helpful for promoting multilingual literacy. This would mean that there is a large inventory of symbols for representing all the available distinctive sounds in all the languages; the appropriate symbols required to write each language would be chosen from this common set. Any particular symbol

would not be used to represent different sounds in different languages (let alone in the same language!).

EXTRACT 2. *Ruhaya Dictionary* (Part I) in Toolbox format

<p>taaguka [ku-tááguka] <i>vi.</i> become torn. <i>chanika</i>. <i>Var:</i> ku-temuka. <i>Prdm:</i> -taagukire.</p> <p>-taagukire [-taagukire] <i>v.</i> torn; tattered. <i>enye kuchanika</i>. <i>Var:</i> -temukire. [<i>Gram:</i> f. ku-taaguka + -ire]</p> <p>taagura [ku-táágura] <i>vt.</i> tear. <i>chana</i>. <i>Var:</i> ku-temura. <i>Prdm:</i> -taagwíre.</p> <p>taaguza [ku-tááguzá] <i>v.</i> walk nakedly. <i>enda uchi</i>. <i>Prdm:</i> -taagwíze.</p> <p>taaha₁ [ku-taaha] <i>vi.</i> come (or go) in; enter; get into; get inside; pierce. <i>ingia; penya</i>. <i>Var:</i> ku-gya; ku-horongotera. <i>Prdm:</i> -tááhire.</p> <p>taaha enju [ku-tááhéju, kutááénju] <i>vp.</i> inaugurate a house or building. <i>zindua nyumba au jengo</i>. [<i>Phon:</i> ku-taaha énju → ku-tááhéju, kutáénju]</p> <p>taaha amagenyi [ku-taahámagenyi] <i>vp.</i> one of the last stages of engagement at which beer and other presents are taken to the bride's family. <i>moja ya hatua za mwisho za kuchumbia ambapo pombe na zawadi nyingine hupelekwa nyumbani kwa msichana</i>. [<i>Phon:</i> ku-taahámagenyi → ku-taahámagenyi → ku-taámagenyi]</p> <p>kitaha-magenyi [kitaha-magenyi] <i>np.</i> a person in charge of the groom's presents to the bride on the send-off day; go between. <i>mjumbe anayesimamia zawadi za bwana harusi kwenda kwa bibi harusi siku ya mkeshá; mshenga</i>. <i>Prdm:</i> -/ba.</p> <p>taaha₂ [ku-taaha] <i>vi.</i> go home. <i>kwenda nyumbani</i>. <i>Prdm:</i> -taahire.</p> <p>-taahire [-tááhire] <i>v.</i> entered; inaugurate (e.g. a new house, building). <i>enye kuíngiwa ndani; enye kuzinduliwa (mf. nyumba au jengo jípya)</i>. [<i>Gram:</i> f. ku-taaha + -ire]</p> <p>-taahize [-tááhize] <i>v.</i> let in; put in. <i>enye kuíngizwa ndani</i>. [<i>Gram:</i> f. ku-taahya + -ire]</p> <p>taahya₁ [ku-taahya] <i>vt.</i> let in; put in. <i>ingiza</i>. <i>Prdm:</i> -tááhize. [<i>Gram:</i> caus., f. ku-taaha]</p> <p>taahya₂ [ku-taahya] <i>vt.</i> take home; put inside the house what has been outside. <i>peleka au ingiza nyumbani vitu vilivyokuwa nje</i>. <i>Prdm:</i> -tááhize. [<i>Gram:</i> caus.]</p> <p>taahya₃ [ku-taahya] <i>v.</i> die (used esp. for chiefs or V.I.P.'s). <i>kufa (hutumika hasa kwa mukama au mtu maarufu)</i>. <i>Prdm:</i> -tááhize. [<i>Gram:</i> ruk., caus.]</p> <p>taana [ku-táána] <i>vi.</i> separate; divorce.</p>	<p><i>tengana; achana</i>. <i>Var:</i> ku-taishukana; ku-taishurana. <i>Prdm:</i> -táine.</p> <p>mihanda etaana [e-mihandeetáána] <i>np.</i> fork in the road; crossroads. <i>njia panda</i>. [<i>Phon:</i> e-mihanda etáána → e-mihandcetáána]</p> <p>taata [tááta] <i>n.</i> my father. <i>baba yangu</i>. <i>Var:</i> tata. <i>Prdm:</i> -/ba.</p> <p>taata-eita-maawe [táátéíta-mááwe] <i>n.</i> my stepfather. <i>baba yangu wa kambo</i>. <i>Prdm:</i> -/ba. [<i>Phon:</i> táátá-eita-mááwe → táátéíta-mááwe]</p> <p>taatento [tááténto] <i>n.</i> my uncle (paternal). <i>baba yangu mdogo</i>. <i>Var:</i> taatento; tatento. <i>Prdm:</i> -/ba.</p> <p>taba [ku-taba] <i>vi.</i> struggle; squirm. <i>hangaika; sumbuka</i>. <i>Prdm:</i> -tabíre.</p> <p>-tabaaliire [-tabááliire] <i>v.</i> [<i>Gram:</i> f. ku-tabaalira + -ire]</p> <p>tabaalira [ku-tabaalira] <i>vt.</i> attack; assault. <i>shambulia</i>. <i>Prdm:</i> -tabááliire.</p> <p>tabaara [ku-tabaara] <i>vi.</i> <i>Prdm:</i> -tabáaire. 1) go hunting; go to war; raid. <i>kwenda kuwinda; kwenda vitani; vamia</i>. 2) (used esp. for chief) make a journey. (<i>hutumika hasa kwa mukama</i>) <i>kwenda safari</i>. <i>Var:</i> ku-zinduka. [<i>Gram:</i> ruk.] 3) die. <i>kufa</i>. <i>Var:</i> ku-taahya; ku-fa.</p> <p>tabaaro [e-tabaaro] <i>n.</i> military expedition; confrontational expedition; voyage. <i>safari ya mapambanoni; safari ya kijasiri</i>. [<i>Gram:</i> f. ku-tabaara]</p> <p>Tabaaro <i>n.</i> name for the year 1968, which means 'Expedition' or 'Voyage'. <i>jina la mwaka 1968, lenye maana ya 'Safari ya mapambanoni'</i>.</p> <p>tabaaruka [ku-tabaaruka] <i>vi.</i> come back from hunting or war. <i>rudi toka kuwinda au vitani</i>. <i>Var:</i> ku-garuka; kw-ihurura. <i>Prdm:</i> -tabáarukire. [<i>Gram:</i> resv., f. ku-tabaara]</p> <p>tabaatirira [ku-tábaatirira] <i>vt.</i> grasp sth (heavy) in hands. <i>fumbata kitu (hasa kizito) mkononi</i>. <i>Prdm:</i> -tabaatiríire.</p> <p>tabaguli [e-tabaguli] <i>n.</i> nonsense; chaos; not well organised; disorganization; disorder; confusion. <i>vurugu; ovyo; isiyo na mpangilio; isiyo na maana</i>. <i>Var:</i> e-tabangu; e-twagu; e-wangaija; e-tabuli.</p> <p>-tabaire [-tabáaire] <i>v.</i> [<i>Gram:</i> f. ku-tabaara + -ire]</p> <p>tabangu [e-tabangu] <i>n.</i> disorganization; disorder; confusion; chaos; nonsense;</p>
--	--

<p>pandemonium. <i>vurugu; ovyo; isiyo na maana; isiyo na mpangilio. Var: e-tabaguli; e-twagu; e-wangaija; e-tabuli.</i></p> <p>tabanguka [ku-tabanguka] <i>vi.</i> mixed up; spoiled; stirred. <i>vurugika; haribika. Prdm: -tabangukire.</i></p> <p>tabangura [ku-tabangura] <i>vt.</i> destroy; spoil; stir; disrupt; mix up. <i>vuruga; haribu; vunia. Var: kw-ita. Prdm: -tabangwire.</i></p> <p>-tabangwire [-tabangwiire] <i>v.</i> [Gram: f. ku-tabangura + -ire]</p> <p>-tabatiriire [-tabatiriire] <i>v.</i> [Gram: f. ku-tabatirira + -ire]</p> <p>-tabiire [-tabiire] <i>v.</i> [Gram: f. ku-tabira + -ire]</p> <p>tabika₁ [ku-tábika] <i>v.</i> vomit. <i>tapika. Var: ku-tanaka. Prdm: -tanakire.</i></p> <p>tabika₂ [ku-tábika] <i>v.</i> branch; bear a branch. <i>ota tawi. Prdm: -tanakire.</i></p> <p>tabira [ku-tabira] <i>vt.</i> bury (generic). <i>zika; fukia na udongo. Var: ku-sangira. Prdm: -tabiire.</i></p> <p>tabuka₁ [ku-tabuka] <i>vi.</i> <i>Prdm: -tabúkire.</i> 1) spring; flow out of the ground. <i>bubujika toka ardhini; toka kama chemchemi.</i> 2) (of winged termites or ants) come out of the</p>	<p>ground. (<i>kumbikumbi au wadudu wengine kutoka ardhini.</i> 3) originate from; come from; emanate from. <i>chimbuka; tokana na.</i></p> <p>tabuka₂ [ku-tábuka] <i>vi.</i> (a branch) break off. (<i>tawi</i>) <i>vunjika. Var: ku-koonyoka. Prdm: -tabukire.</i></p> <p>tabuli [e-tabuli] <i>n.</i> chaos; disorganization; disorder; confusion. <i>vurugu; ovyo; isiyo na mpangilio. Var: e-tabaguli; e-tabangu; e-twagu; e-wangaija.</i></p> <p>tabura₁ [ku-tabura] <i>vi.</i> boil up; bubble up. <i>chemka; tutuma. Prdm: -tabwire.</i></p> <p>tabura₂ [ku-tábura] <i>vt.</i> break off (a branch). <i>vunja (tawi). Prdm: -tabwiire. [Gram: resv., f. ku-tabika]</i></p> <p>-tabwire₁ [-tabwiire] <i>v.</i> [Gram: f. ku-tabura + -ire]</p> <p>-tabwire₂ [-tabwiire] <i>v.</i> broken off (e.g. a branch). <i>enye kuvunjwa (mf. tawi). [Gram: f. ku-tábura + -ire]</i></p> <p>tagaarura [ku-táagarura] <i>vt.</i> make a very loud noise. <i>toa mlio mkali. Prdm: -tagaarwire.</i></p> <p>tagaarura akanwa [ku-táagaruráakanwa] <i>vp.</i> scream; yell. <i>lia kwa sauti; lia vikali. Var: ku-borooga; ku-kunga; ku-chura.</i></p>
--	---

Where: *appl* = applicative; *caus* = causative; *f* = from; *gram* = some grammatical information; *morph* = morphological information; *n* = noun; *phon* = phonological information; *prdm* = paradigm (i.e. *-ire* form for verbs, and noun class for nouns); *rev* = reverse; *ruk* = palace register; *v* = verb; *vi* = intransitive verb; *vt* = transitive verb; *var* = variant; *vp* = verb phrase; *Zb* = Ruziba dialect.

A number of obstacles lie in the way of achieving this ideal. Foremost is the absence of a central authority to promote such a standard and to organize orthography conferences to develop it. If ECLs were permitted in the public domain, the need to develop materials in the ECLs would be a compelling reason to support the creation of the uniform set of conventions; the prohibition on ECLs in the schools and mass media removes the most immediate inspiration for this work. An equally formidable obstacle is the force of established traditions, given that any existing written materials evolved mainly around particular missionary endeavours (see Endl and Thomson 2002). These were at times in hostile competition with each other, and used different conventions for the same or related languages.

The influence from Kiswahili is particularly problematic. For most people this is the only language they ever learn to read (with any degree of success). They are never taught to read or write the ECL mother tongue. So when presented with an opportunity to read or write something in the ECL, the conventions they rightly fall back on are from Kiswahili. That is fine if the ECL is close to Kiswahili. But many ECLs will have significant differ-

ences, considering that all four major African language families—Niger-Congo, Nilo-Saharan, Afro-Asiatic, and Khoisan—are represented in Tanzania, as illustrated below.

1. Niger-Congo—Benue-Congo—Bantoid—Bantu—*Kiswahili, Sukuma*
2. Afro-Asiatic—Cushitic—*Iraqw, Ma'a*
3. Nilo-Saharan—Nilotic—*Maasai, Luo*
4. Khoisan—*Sandawe, Hadza*

One such difference is in the vowel system—whether there are five or seven (with long and short ones). So although Sukuma (F21 in Guthrie's classification [1968, 1967–1971]) has seven vowels, the earliest Sukuma Bible (Bible Society of Tanzania 1960) uses five vowels as in Kiswahili.² Another difference is in the consonant inventory: the non-Bantu languages, in particular, have many sounds not found in Kiswahili, including clicks in Sandawe (a Khoisan language),³ and uvulars and pharyngeals in Iraqw (an Afro-Asiatic language). For instance, Iraqw orthography includes two letters not used in writing Kiswahili, *q* for the voiceless uvular stop, and *x* for the voiceless velar fricative. It also uses symbols that are not even part of the Roman alphabet, including a slash / for the pharyngeal fricative, and an apostrophe ' for the glottal stop (Mous et al. 2002). Symbols like these present particular challenges to even the most talented ECL literacy lover who has not received basic guidance.

Even within the project it has not been possible to avoid these problems and promote a common set of conventions. For instance, the Runyambo (Rugemalira 2002), Kihangaza (Rubagumya 2006), and Ciruuri (Massamba 2005) lexicons use *c* instead of the Kiswahili *ch*. But the Ruhaya dictionary (Muzale 2006), with a strong tradition of using *ch* in the orthography, sticks to that symbol. Another example concerns the discrepancies in abiding by the phonemic principle in the adopted orthographies: in both Runyambo and Ciruuri [b] and [β] are allophones—the stop only appearing after the bilabial nasal. The Runyambo lexicon uses one orthographic symbol *b* for the phoneme (and has been criticized by some users for that). The Ciruuri lexicon uses two symbols, *b* and *bh*. Again, although no phonemic distinction exists between [r] and [l] in Ciruuri, the author of the lexicon retains both orthographic symbols *r* and *l* interchangeably.

A similar observation applies to the representation of vowel length. In Kihangaza, Ruhaya, Ciruuri, and Runyambo, vowel length is distinctive, with five short and five long vowels. Yet in some phonetic contexts vowels are predictably lengthened—particularly after a consonant + glide sequence and before a nasal consonant + stop sequence. The published lexicons for these languages show the tension that exists in attempting to estab-

² The latest New Testament translation uses seven vowels (Bible Society of Tanzania 2000b)

³ Kagaya (1993:vii) identifies twelve clicks in Sandawe, whereas Eaton (2005:12) has fifteen after making a distinction between plain and voiced clicks.

lish an orthography that makes a phonemic representation but also reflects the way people speak. The Ciruuri lexicon in particular, marks both phonemic long vowels and predictably long vowels by doubling them. In the other lexicons the departures from the phonemic principle appear to be attributable to editorial oversight rather than conscious decision.

The representation of tone is another example of a compromise of the desire to capture in writing the way people speak. Although the agreed-upon ideal within the project is to mark tone on the entry word, in actual practice this has not been easy to achieve. So some of the lexicons have no tone marking, or they have only partial tone marking. Even where it would be possible for a researcher to mark tone on all ECL materials in the lexicon (or even in a narrative text), it may not be desirable to do so because the text might appear too complex for the ordinary user. Such a scenario is the equivalent of a phonetic representation at the segmental level.

3.2 GRAMMAR. The production of descriptive grammars has evolved much more slowly than that of lexicons. It would appear that this kind of documentation is more restricted in audience and more demanding on the researcher. So far only one such grammar—for Runyambo—has been published (Rugemalira 2005a). Grammar notes exist for three other languages—Kihangaza, Kimashami, and Cigogo.

The basic tool for collecting morpho-syntactic data is the 256-sentence list originally developed by Herman Batibo (ca. 1990). The sentences were intended to elicit Bantu nominal prefixes, verb forms, and basic sentence structure. They enable the researcher to arrive at an outline of the morphology and syntax of a Bantu language. The Runyambo grammar developed along these lines has a general introduction, and a chapter each on phonology, nominal morphology, verb morphology, and phrase structure. The targeted reader is clearly the linguist.

If the project were to target a different type of audience, it would be necessary to change the form of the envisaged grammar, particularly the language of discourse. Kiswahili would be the best choice in this regard, since the metalanguage is much more developed in comparison with that of the ECLs. Eventually, it will be important for some of these languages to produce grammars written in the native languages themselves. Needless to say, the realization of this ideal will require significant changes in the national policies and practices pertaining to language use. Such changes might include the creation of some room in the school curriculum for teaching literacy in ECLs, as well as permission to use ECLs in the media—which would create an instant market for advanced skills in writing and speaking certain ECLs. In this connection, the Ruhaya and Ciruuri lexicons have taken the step of including Ruhaya and Ciruuri versions of their respective introductions. This type of effort makes a significant contribution to the development of the requisite metalanguage.

3.3 ORAL LITERATURES. Although the original project objectives did not include the documentation of oral literatures, such as stories, folk tales, proverbs and riddles, poetry, and song, arrangements have already been made to incorporate this objective. So far, three collections of folk tales have been published under the auspices of the project (see Appendix 1).

4. THE TANZANIA LANGUAGE ATLAS.

4.1 IDEAL AND REALITY. The objective of producing a language atlas of Tanzania has presented formidable challenges. The ideal language atlas would be a very informative resource that would be of interest to a wide range of users. It would name all the languages spoken in the country, showing the number of people who speak each as a first or second language, the geographical distribution of the speakers, and various indicators of each language's vitality (e.g., domains of use, acquisition by children, attitudes by users and neighbours, and available literature).

One of the most cost-effective ways to obtain part of the requisite information would be to include a set of questions in the national population census. Even in this case, only a few question items would be allowed, and the linguist would have to demonstrate the need for such questions to the demographers and politicians in the census bureau. And given the chance to frame questions on languages, the utmost care would have to be taken, since the possible pitfalls are legion. Gadelii (2001) has shown, in the case of Mozambique, that formulating and interpreting a language question in a census can be quite complex. The main issues revolve around the following questions. What is it to know/speak a language and how can this knowledge be authenticated for self-reporting respondents in a census? How can a person's first language be distinguished from his/her second language in a multilingual community? How is a language to be distinguished from a dialect? How can respondents be made to distinguish the language they speak from their ethnic identity?

As it turned out, it was not possible to include language questions in the 2002 national population census questionnaires, because it was considered not politically acceptable. The Census Commissioner took the opportunity to make the point that Tanzania is past the stage of counting tribes and has made giant strides towards the creation of a homogeneous nation with one national language (Damas Mbogoro, personal communication). He maintained that any activity making reference to tribal languages is retrogressive in that regard.

The Census Commissioner's stance needs to be viewed within the wider context of government policies and practices. Although the following specific, albeit isolated, government statement on ECLs is fairly impressive, in effect there have been more potent forces working against the promotion of these languages. The statement is contained in *Sera ya Utamaduni* [Cultural Policy] (Tanzania Government 1997:17–18) and affirms the need:

1. to promote research, preservation, and translation of the ethnic community languages;
2. to produce dictionaries and grammars of the ethnic community languages;
3. to publish various materials in the ethnic community languages.

Two counter-forces may be identified. The first force working against the ECLs has "national unity" as its war cry. It is argued that the promotion of the multiple ECLs would be inimical to national unity (Mkude 2002). In this regard the parallel between "one country–one language" policies with "one party democracy" is instructive. But while the freedom of association supported by the current multi-party situation may grow and become

strong with time, the freedom of thought and expression inherent in multilingualism may be suffering irreversible setbacks under the current policies of linguistic minimalism.

The second force working against ECLs is the association of specific languages with modernity and progress. The failure to resolve the issue of the language of instruction in schools captures the malaise. Just as it is argued that English is the language of science and technology and therefore the appropriate medium for secondary and further education, similarly, Kiswahili is regarded as the appropriate medium for primary education, it being the national lingua franca that takes an individual away from parochial domestic and tribal concerns.⁴ In either case, education in the ECLs is viewed as detrimental to the national goal of progress.

4.2 THE EVOLUTION OF A METHOD. In a series of trial studies beginning in 2001, the Languages of Tanzania Project refined a method for obtaining information on the languages spoken in the country. In the first research tour, four researchers, with two assistants each, began to assemble information on the languages spoken in three districts of Kagera Region—Karagwe (one researcher), Muleba (one researcher), and Bukoba (two researchers). The researchers determined, from speaking to people in these areas, that although there were two major languages (Ruhaya and Runyambo) spoken in the region, there were also a few other languages spoken by minorities (Mreta et al. 2002).

In 2002, the project compiled a list of informants—mostly students at the University of Dar es Salaam—who could provide information on the languages of the districts in the Lake Victoria zone. These were people who had been born in the respective areas, who spoke the languages of those areas, and who were otherwise knowledgeable about the general linguistic situation of the areas. These informants were asked to provide information on the two major languages spoken in their districts, with a percentage indication of their proportional strength. Although attempts were also made to obtain details on dialectal variations, this information proved to be too complex to process.

On the basis of this data on languages spoken in each district, it was possible to calculate the number of speakers for each of the major languages identified, using the 1988 population census statistics. A map capturing this information was eventually produced covering the five regions—Kagera, Mwanza, Mara, Kigoma, and Shinyanga. Several shortcomings became apparent. First, the census data were already fourteen years old; in a country where there is a high rate of mobility, the information could hardly be taken to represent the existing situation. Second, the data collection procedure had assumed that in each ward (sub-division of a district) there would be only two major languages; any other languages were ignored. The result is that the methodology failed to capture languages that are spoken by smaller proportions of the community. Third, it was realized that there were many “languages” in the final list that could arguably be regarded as dialects of some bigger language; in the final analysis, an editorial decision needs to be made on such distinctions. Fourth, the map that was eventually printed highlighted the administrative

⁴ For the perennial debate on the language of instruction in Tanzania, see, among others, Rubagumya 1990; Roy-Campbell and Qorro 1997; Brock-Utne, Desai, and Qorro 2004, and the various references cited therein. The journal of the Faculty of Education of the University of Dar es Salaam published a special edition on language of instruction in 2006.

boundaries and, in some cases, language boundaries, which obscured the language names and failed to consistently indicate all language boundaries. Finally, it was noted that the information available was not sufficient to produce a language atlas; and also, in particular, that it was important to capture some sociolinguistic information.

A pilot study done in the Arusha Region in early 2005 was designed to address the paucity of sociolinguistic information. The questionnaire was detailed, covering the various domains of use for the ECLs and seeking to determine the vitality levels for the languages in question (Languages of Tanzania Project 2005). The first six questions sought information on the first/mother tongue of the informant and its status in the village (whether native or immigrant, minority or majority) as well as other languages spoken by the informant. Questions seven to fourteen asked about the patterns of migration in and out of the village and the associated language behaviours. Questions fifteen to eighteen sought to determine the linguistic profile of the family— the languages of the informant's spouse, children, and parents, as well as the patterns of language use between the informant and these relatives. Questions nineteen to twenty-eight dealt with available types of literature in the informant's language and language use in worship, in the marketplace, in letters to various people (mainly relatives), in various government offices and by various officials at the village and district levels. One hypothetical question asked what the preferred language would be if a community radio were to be established.

The experience of the pilot study made it clear that a single researcher with six assistants could hardly be expected to do justice to the demands of this questionnaire in a logistically challenging region like Arusha in the four weeks available. The lesson from that study was that a carefully controlled research project on the sociolinguistic profile for each of the languages of Tanzania would require more time and resources than are currently available to the project.⁵

4.3 ATLAS DATA COLLECTION. The final atlas that is being compiled is based on data obtained in three stages. First, the 2002 population statistics for each village in Tanzania were obtained from the National Bureau of Statistics (Census Office); the information was edited and stored in the project's data base using both Excel and Access formats. Second, a list of potential informants was compiled: this consisted mainly of students at the University of Dar es Salaam who spoke the various languages and grew up in the various regions where these languages are mainly spoken. In a series of workshops in 2005–2006, each lasting one day, these informants provided information on the various languages (up to five) spoken in each village (for rural areas) and street (for urban areas) shown in the population census database. The informants were required to indicate what languages, up to a maximum of five, are found in each village or street (as mother tongue) and to estimate, in percentage points, the relative number of speakers of each language. Using these estimates

⁵ Direct funding for the project has been provided by the Swedish Agency for Research Cooperation (SAREC), the research department of the Swedish International Development Agency (SIDA), as part of research support to the University of Dar es Salaam. The project has relied on the goodwill and interest of the linguists in the language units of the university, namely, the Department of Foreign Languages and Linguistics, the Department of Kiswahili, and the Institute of Kiswahili Research. Needless to say, these experts could only squeeze in work for the project in their already overloaded briefs if and when feasible.

and the population data, it has been possible to determine the number of speakers for each language in each village/street. The appropriate aggregations make it possible to produce data for any administrative unit above the village/street (i.e., ward, district, region, up to the national level).

Third, project researchers spent six weeks in July–August 2006 in all the regions of Tanzania filling gaps where no information was obtained during the university campus workshops, assessing the validity of the data by cross-checking with other people in the regions, and correcting errors, particularly where percentage points for a street/village's languages were in excess of or less than 100.

The most problematic areas to cover with this method of identifying languages and their speakers were the large urban areas of the city of Dar es Salaam and parts of the city of Mwanza (especially Nyamagana District). In these highly metropolitan areas, people (the parents, at least) in every other house will speak a different language. But it is in these areas that Kiswahili is likely to be the first language of the young generation, so that the statistical deficit ought to be credited to Kiswahili. People born in and/or raised in Dar es Salaam during the last thirty years are likely to have Kiswahili as their first language, and the population under thirty years old accounts for seventy percent of the population of Dar es Salaam. This amounts to 1.7 million out of the 2.4 million people reported for this city in the 2002 census. By contrast, in the smaller regional towns it is still largely possible to identify concentrations of speakers of the local languages with reasonable accuracy.

4.4 PRELIMINARY RESULTS. Analysis of the language atlas data is in the final stages, and the atlas is due to be completed by the end of the year. Although a few more editorial decisions on the form and content of the atlas will be made in the coming months, it is unlikely that such decisions will alter the current results substantially.

TABLE 1: Languages of Tanzania

1. Sukuma	5,194,553	21. Nilamba	385,824	41. Pogoro	200,974
2. Swahili	2,361,170	22. Mwera	385,408	42. Ndamba	196,179
3. Ha	1,228,864	23. Rangi	370,578	43. Mashami	194,868
4. Gogo	1,023,618	24. Jita	364,889	44. Ndali	192,537
5. Nyamwezi	958,898	25. Sumbwa	361,111	45. Zinza	187,105
6. Haya	828,272	26. Nyambo	355,975	46. Luo	185,172
7. Makonde	803,974	27. Kagulu	336,749	47. Ngindo	170,329
8. Maasai	803,463	28. Makua	300,492	48. Rombo	167,478
9. Hehe	740,113	29. Safwa	299,908	49. Digo	166,347
10. Nyakyusa	732,990	30. Nyiha	275,864	50. Hangaza	155,421
11. Fipa	712,786	31. Chagga	273,474	51. Kwere	151,568
12. Iraqw	602,661	32. Matengo	270,778	52. Vunjo	141,853
13. Bena	592,370	33. Zalamo	259,511	53. Nyamwanga	141,281
14. Sambaa	565,276	34. Ngoni	254,699	54. Ndengereko	139,224
15. Nyaturu	552,344	35. Rundi	246,910	55. Datooga	138,777
16. Asu	529,455	36. Kerewe	240,941	56. Ndendeule	135,753
17. Zigua	436,762	37. Matumbi	226,791	57. Subi	135,479
18. Kurya	423,511	38. Kinga	217,173	58. Bondei	121,934
19. Yao	416,802	39. Meru	217,115	59. Sangu	119,342
20. Luguru	403,455	40. Nguu	214,533	60. Leki	114,990

61. Arusha	113,921	94. Burunge	27,942	127. Hadzabe	6,289
62. Gorwaa	112,941	95. Mpoto	27,559	128. Somali	6,063
63. Sagara	106,331	96. Kutu	27,509	129. Nindi	5,689
64. Maraba	97,601	97. Ganda	26,343	130. Kami	5,518
65. Zanaki	97,429	98. Nyisanzu	25,978	131. Rooba	5,416
66. Pangwa	95,129	99. Ruuri	25,924	132. Lomwe	5,203
67. Shubi	85,128	100. Vidunda	25,318	133. Mambwe	5,056
68. Nyarwanda	81,635	101. Gweno	25,234	134. Shashi	4,449
69. Woso	81,181	102. Sonjo	24,618	135. Surwa	4,394
70. Malila	77,934	103. Segeju	23,232	136. Ndwewe	4,358
71. Kwaya	74,153	104. Taturu	22,672	137. Kwiva	4,216
72. Sandawe	65,935	105. Kisi	22,395	138. Wemba	3,908
73. Pimbwe	64,592	106. Songwe	21,980	139. Kwavi	2,737
74. Kimbu	62,672	107. Bembe	21,915	140. Ndonde	2,458
75. Nyambwa	60,390	108. Matambwe	21,343	141. Gala	2,212
76. Alagwa	52,816	109. Kenye	19,397	142. Burushi	1,887
77. Ngoreme	52,360	110. Ikoma	19,393	143. Lungu	1,484
78. Ikizu	48,456	111. Longo	18,253	144. Gusii	1,468
79. Mochi	47,144	112. Kabhwa	17,692	145. Nkamanga	1,396
80. Nyasa	45,214	113. Rieri	14,246	146. Ndorobo	1,152
81. Manda	43,115	114. Nyankore	12,876	147. Bwali	1,070
82. Wanji	41,800	115. Kongo	11,916	148. Kiga	662
83. Bende	41,490	116. Kamba	11,790	149. Hanju	654
84. Uru	40,364	117. Mbunga	11,589	150. Magindo	474
85. Simbiti	38,086	118. Sweta	10,735	151. Kikuyu	307
86. Tongwe	37,686	119. Kine	9,437	152. Markanani	264
87. Mbugwe	37,177	120. Kahe	9,130	153. Lingala	237
88. Mengwe	34,747	121. Isenye	8,238	154. Manyema	159
89. Ma'a	33,653	122. Doe	7,944	155. Tonga	141
90. Nyagatwa	33,434	123. Tiliko	7,171	156. Wanda	136
91. Pokomo	32,111	124. Nata	7,050		
92. Lambya	30,408	125. Hacha	7,008		
93. Bungu	30,332	126. Rwingo	6,451		

Table 1 shows the full list of 156 languages spoken in Tanzania with their number of speakers. As the table shows, the language with the greatest number of speakers by far is Sukuma, with some 5,000,000 speakers, followed by Kiswahili and Ha. The language with the smallest number of speakers is Wanda, with only 136 speakers. It will be noted that the top ten languages account for 14,671,313 speakers, which is 46% of the total. The bottom 50 languages account for 309,000 speakers only, which is about 1% of the total population. This gives an indication of the serious language endangerment situation in the country, although some of these small languages are fragments of larger speech communities across the borders.

Language names that do not appear in this table may have failed to make it for a number of reasons. One reason may be that the name not appearing here is deemed to be an alternative name, particularly one used by neighbors/enemies/outsideers but not acceptable to the speakers themselves. An example is the term Mang'ati (pronounced [maŋati]) to refer to Datooga (or Barbaig). The former term is used by neighbors and means "man eaters." A second reason may be that the language name is deemed to refer to only a dialect of

TABLE 2: A sample of language distribution across regions

Language	Arusha	Dodoma	Kagera	Kigoma	Mbeya	Morogoro	Mwanza	Shinyanga	Tabora	Tanga
Sukuma	1,522	2,364	201,244	-	24,790	7,586	1,914,618	2,232,027	616,747	810
Swahili	60,715	143,926	35,952	516,076	3,459	125,519	401	91,936	7,756	102,995
Ha	-	-	83,613	936,915	184	1,003	39,473	38,436	105,336	508
Gogo	30	818,583	-	-	2,223	38,178	15	-	588	954
Nyamwezi	-	-	-	894	6,935	2,011	5,614	82,032	795,405	37
Haya	888	733	798,370	39	658	193	20,310	2,908	3,895	-
Maasai	515,671	12,628	-	-	6,733	47,865	-	418	-	22,278
Hehe	65	50,973	-	-	16,962	69,861	-	-	-	168
Nyakvusa	151	-	-	-	682,508	20,815	-	-	4,851	-
Fipa	-	-	-	5,884	12,411	-	6	-	12,124	245
Iraqw	154,164	311	-	-	-	158	-	267	-	505
Bena	-	20,156	-	-	33,943	15,451	-	-	-	18,960
Sambaa	7,469	45	-	-	-	1,865	-	131	-	535,205
Nyaturu	10,624	12,243	-	-	-	-	1,522	4,875	2,178	-
Asu	30,501	112	-	-	-	10,665	-	-	-	102,320
Zigua	-	2,033	-	-	-	7,631	-	-	-	381,068
Kurya	13	84	-	-	-	569	14,221	5,485	632	43
Yao	-	-	-	-	2,215	1,046	-	-	-	205
Luguru	-	-	-	-	-	402,524	-	-	-	-
Nilamba	5,213	-	-	-	221	-	1,308	39,336	44,584	-
Rangri	33,118	282,062	-	390	-	-	-	33	4	-
Jita	-	-	20,096	-	-	172	136,942	12,965	348	-
Sumbwa	-	-	4,306	-	-	-	79,830	239,863	37,112	-
Nyambo	-	-	355,975	-	-	-	-	-	-	-
Kagulu	-	102,301	-	-	-	231,837	-	-	-	1,232
Safwa	932	-	-	-	290,404	967	-	-	-	-
Nyiha	-	-	-	-	274,431	-	-	-	-	-
Chagga	97,785	5,792	-	-	2,613	22,867	5,256	7,444	8,007	12,515
Zalamo	-	-	-	-	-	2,097	-	-	-	-
Ngoni	-	-	-	-	3,084	6,462	-	156	-	5,107
Rundi	-	-	17,882	52,070	-	-	-	937	32,164	2,214
Kerewe	-	-	20,212	-	-	-	195,685	-	-	11
Matumbi	-	-	-	-	-	834	-	-	-	112
Kinga	-	-	-	-	43,386	1,665	-	-	-	61
Meru	205,860	-	-	-	-	-	-	-	-	-
Nguu	-	16,005	-	-	-	95,622	-	-	-	80,270
Pogoro	-	69	-	-	-	196,876	-	-	-	-
Ndamba	-	-	-	-	-	195,995	-	-	-	-
Mashami	126	-	-	-	-	-	-	-	-	167
Ndali	-	-	-	-	187,136	633	-	-	561	-
Zinza	-	-	16,043	349	-	-	169,955	635	-	-
Luo	-	-	225	-	-	-	9,602	2,555	101	-
Ngindo	-	-	-	-	-	32,995	-	-	-	-
Digo	-	-	-	-	-	-	-	-	-	166,234
Hangaza	-	-	155,421	-	-	-	-	-	-	-

another language which is listed. For instance, Dzungwa is a dialect of Hehe; Jinakiya and Dakama are dialects of Sukuma; Tumbatu is a dialect of Swahili; and Ziba is a dialect of Haya. Finally, a language name may be missing from our list because the language was too small to be among the five major languages of the village/street. At the village level, a sixth or seventh language would be scarcely significant in statistical terms. And this leads back to the earlier discussion about the major metropolitan areas: none of the Asian immigrant community languages (e.g., Gujarati, Arabic, Hindi) surfaces in the list because these are to be found primarily in the major urban areas of Dar es Salaam and Mwanza. But these areas, as already indicated, could not be handled with the project's estimates procedure.

TABLE 3. Language distribution by number of districts

1. Swahili	98	34. Manda	8	107. Kami	2
2. Chagga	63	35. Zalamo	8	108. Kenye	2
3. Sukuma	56	36. Digo	7	109. Kwavi	2
4. Maasai	36	37. Kamba	7	110. Lingala	2
5. Ha	34	38. Malila	7	111. Ma'a	2
6. Nyakyusa	33	39. Ndamba	7	112. Magindo	2
7. Hehe	32	40. Subi	7	113. Matambwe	2
8. Nyamwezi	32	41. Kimbu	6	114. Mbugwe	2
9. Asu	31	42. Ndendeule	6	115. Mbunga	2
10. Haya	31	43. Nyambo	6	116. Nata	2
11. Gogo	29	44. Nyamwanga	6	117. Ndonde	2
12. Bena	28	45. Sandawe	6	118. Ndorobo	2
13. Makonde	26	46. Zanaki	6	119. Ndwewe	2
14. Kurya	25	47. Mashami	5	120. Nyankore	2
15. Nilamba	25	48. vunjo	5	121. Nyisanzu	2
16. Ngoni	24	49. Gweno	5	122. Pimbwe	2
17. Nyaturu	24	50. Hadzabe	5	123. Rooba	2
18. Sambia	22	51. Kisi	5	124. Ruuri	2
19. Yao	22	52. Kwaya	5	125. Segeju	2
20. Jita	21	53. Ndengereko	5	126. Shashi	2
21. Kinga	21	54. Bende	4	127. Shubi	2
22. Datooga	20	55. Bungu	4	128. Sweta	2
23. Rangi	17	56. Rombo	4	129. Temba	2
24. Iraqw	16	57. Ikizu	4	130. Gala	1
25. Luo	16	58. Ikoma	4	131. Ganda	1
26. Safwa	16	59. Mambwe	4	132. Hacha	1
27. Fipa	15	60. Matengo	4	133. Hanju	1
28. Rundi	15	61. Simbiti	4	134. Kahe	1
29. Zigua	15	62. Alagwa	3	135. Kiga	1
30. Kerewe	14	63. Arusha	3	136. Kikuyu	1
31. Makua	14	64. Bembe	3	137. Kine	1
32. Mwera	14	65. Burunge	3	138. Kongo	1
33. Nyarwanda	14	66. Woso	3	139. Kwiva	1
34. Matumbi	12	67. Doe	3	140. Lomwe	1
35. Nguu	12	68. Gorwaa	3	141. Lungu	1
36. Taturu	12	69. Hangaza	3	142. Manyema	1
37. Zinza	12	70. Kutu	3	143. Maraba	1
38. Meru	11	71. Lambya	3	144. Markanani	1
39. Ndali	11	72. Longo	3	145. Mengwe	1
40. Ngindo	11	73. Ngoreme	3	146. Mpoto	1
41. Pogoro	11	74. Nindi	3	147. Nkamanga	1
42. Sumbwa	11	75. Nyagatwa	3	148. Nyambwa	1
43. Kagulu	10	76. Sagara	3	149. Pokomo	1
44. Nyiha	10	77. Sonjo	3	150. Rieri	1
45. Pangwa	10	78. Tongwe	3	151. Rwingo	1
46. Sangu	10	79. Vidunda	3	152. Songwe	1
47. Luguru	9	80. Burushi	2	153. Surwa	1
48. Nyasa	9	81. Bwali	2	154. Tiliko	1
49. Somali	9	82. Mochi	2	155. Tonga	1
50. Wanji	9	83. Uru	2	156. Wanda	1
51. Bondei	8	84. Gusii	2		
52. Kwere	8	85. Isenye	2		
53. Leki	8	86. Kabhwa	2		

The treatment of the Chagga cluster of languages is somewhat peculiar. In the Chagga homeland at the foot of Mount Kilimanjaro, it was possible to identify six languages, namely Mashami, Mochi, Vunjo, Woso, Rombo, and Uru, spoken in distinct areas. Outside this homeland, however, short of doing a house to house count, it is difficult to determine which of these six languages is spoken by a person from Chaggaland.⁶ This is probably one of the best examples of a single ethnic identity not matched by one language.

Table 2 presents the patterns of geographical distribution for a sample of languages across ten regions (out of the twenty regions of the mainland, i.e., excluding Zanzibar).⁷ Table 3 presents the distribution of each language across districts. The numbers in Table 3 indicate the total number of districts in which the language is significantly located. As the tables show, some languages are represented substantially in many regions and districts, while other languages are more local, being restricted to the original homeland of the speakers. Apart from Kiswahili (found in 98 out of the 125 districts), Chagga is by far the most widely distributed ECL geographically; it is found in 63 districts across the country, closely followed by Sukuma in 56 districts. The two languages can be regarded as close rivals to the national language, Kiswahili, in this regard. It may be observed that in the top league, number of speakers does not match geographical spread significantly.

4.5 SIGNIFICANCE OF THE RESULTS. These results are probably the best estimates for the languages of Tanzania as a whole (cf. Grimes 2000). Yet it is important to note the areas of possible weakness and gaps in these results. First, the estimates are based on the official census statistics; any weaknesses in the base data will be reflected in our statistics of speakers of the various languages. Second, the results are based on a distinction between a first/mother language and a second language. There are many areas where Kiswahili is the major language, but it is not the mother tongue of the people who use it; it is a second language. What these results show are the major languages that are mother tongues of the people that were counted. If a person spoke two or more languages, that was irrelevant for our purposes. We assumed that every person had only one first/mother tongue.

The concept of “mother tongue/first language” (L1) is by no means simple, nor is this definition of ours universally shared: the language first (chronologically) acquired from the parents or caretakers and not necessarily the language spoken by the biological mother—there being a significant difference in the manner in which any subsequent language is added to one’s repertoire. There may be cases where Kiswahili or another major community language overtakes the mother tongue in domains and frequency of use, so that fluency in the mother tongue declines or atrophies, at least in many domains. It may be that a person grows up in a bilingual community and acquires two languages simultaneously; or that in such a community the child predominantly acquires the language of the wider community and has only receptive skills in the language of the home. A related case, which

⁶ The total of all Chagga people amounts to 946,362, so that if it is regarded as one language Chagga ranks sixth after Nyamwezi (958,898). But it may be noted that the linguistic difference between Mashami and Rombo, for instance, is probably greater than the difference between Sukuma and Nyamwezi.

⁷ The whole population of Zanzibar is counted as Kiswahili speaking.

the linguist can dismiss, is that of a person who regards the ethnicity as dictating the mother tongue, in the same way that, in a patrilineal system, the clan (and totem) of the child follows from that of the father.

A third weakness in these results is that informants were more likely to be thinking about the heads of the households under investigation and to ignore any special circumstances regarding the languages of the children. So even in the smaller regional towns there may have been some undercounting of Kiswahili-speaking young people whose parents are identified with a different mother tongue. A fourth problem is that it is not easy to provide a list of the languages that have died in the last thirty to fifty or so years. This would require the existence of an earlier inventory of languages similar to the current one in the way languages are identified. Then it would be possible to say that language *Y* in the old list does not appear on the current list and so is assumed to be dead. A related issue stems from the way in which these results were obtained, which may not have noted certain critically endangered languages because they had too few speakers—that is, they were not in the top five languages of any area.⁸ Finally, in the absence of detailed linguistic studies of the languages in question, the distinction between language and dialect has been based on the information available to the research team, with clear recognition of the fact that such a distinction may not be entirely or even primarily a linguistic matter.

5. THE CHALLENGES AHEAD. The level of language endangerment for even the most populous speech communities is considerable given the ever-rising fortunes of Kiswahili.⁹ This poses two related problems: first, the vocabulary, grammar, and knowledge preserved in the relevant language may be eclipsed and disappear without a trace since there are no records. Second, the communities caught in the midst of this transition may be rendered powerless, without the ability to participate in the decisions that shape their lives and their future, on account of their language not being the language of power. If members of a community cannot use their language to debate the policies affecting them, or to question the laws that are enacted to control them, these communities will effectively be marginalized, even when their grandchildren have fully adopted the national or global language.

For these reasons, it is important to carry on language documentation and to study the sociolinguistic profiles of the ECLs. For the Languages of Tanzania Project, it will be necessary to incorporate an advocacy component as a core activity of the project, in order to make a case for the legal opening up of the public domain for the ECLs. These languages need to be allowed space in the mass media, in the schools, and in political discourse. Advocates need to realize the formidable obstacles ahead. They need to allay

⁸ Speakers of minority or dying languages are likely to associate themselves with larger and more prestigious languages; they are more likely to be ignored by a nonlinguist informant and get lumped into the larger community language pool.

⁹ There is a sense in which Kiswahili is itself under pressure from English. It is not being given full room to exercise its potential in the various public domains—in education, government bureaucracy, and commerce. Even at the primary school level, its status as the language of instruction appears to be in danger, not just from the popularity of the private English-medium schools (Rugemalira 2005b), but from official policies as well: in Zanzibar it has already been decided to revert to English as the language of primary education (Rubagumya, personal communication).

politicians' fears of the dangers of tribalism vis-à-vis national unity, but also to confront their fears of democracy and self-determination of communities. Activists need to make a case to the society in general, and planners in particular, for the need to spend resources on "antiquated" or "dying" languages in the face of competing demands. They need to impress upon all those in charge the role of culture in creating and recognizing one's self worth and confidence as the springboard for participating and competing against other people in the global economy.

APPENDIX 1: List of publications produced by the Languages of Tanzania Project.

	Author	Title	Publisher	Year
1	J. Rugemalira	Runyambo Lexicon	LOT Project, UDSM	2002
2	J. Maho & B. Sands	The Languages of Tanzania: A Bibliography	Goteborg: Acta Universitatis Gothoburgensis	2002
3	H. Muzale	Tanzanian Sign Language Dicionary	LOT Project, UDSM	2004
4	D. Massamba	Eciruuri Lexicon	LOT Project , UDSM	2005
5	J. Rugemalira	A Grammar of Runyambo	LOT Project , UDSM	2005
6	LOT Project	Occasional Papers in Linguistics 1	LOT Project , UDSM	2005
7	C. Rubagumya	Igihangaza Lexicon	LOT Project , UDSM	2006
8	H. Muzale	Ruhaya Dictionary	LOT Project , UDSM	2006
9	K. Legere & P. Mkwan'hembo	Hadithi za Kividunda 1	TUKI, UDSM	2006
10	K. Legère	Ngh'wele-Swahili- English Wordlist	Gothenburg University	2006
11	J. Rugemalira	Kimashami Dictionary	LOT Project	2008
12	Y. Rubanza	Luzinza Lexicon	LOT Project	2008
13	A. Mreta	Kisimbiti Lexicon	LOT Project	2008
14	S. Sewangi	Kiikizo Lexicon	LOT Project	2008
15	J. Mdee	Kijita Lexicon	LOT Project	2008
16	K. Kahigi	Sisumbwa Dictionary	LOT Project	2008
17	A. Mreta	Chasu Lexicon	LOT Project	2008
18	Y. Rubanza	Kimeru Lexicon	LOT Project	2008
19	S. Sewangi	Kigweno Lexicon	LOT Project	2008
20	G. Mrikaria	Kimochi Lexicon	LOT Project	2008
21	K. Kahigi	Kikahe Lexicon	LOT Project	2008
22	J. Maghway	Gorwaa Lexicon	LOT Project	2008
23	J. Kiango	Kibondei Lexicon	LOT Project	2008
24	Z. Mochiwa	Kizigula Lexicon	LOT Project	2008
25	LOT Project	Occasional Papers in Linguistics 2	LOT Project	2008

26	LOT Project	Occasional Papers in Linguistics 3	LOT Project	2008
27	M. Petzell	Kagulu: Grammar, Vocabulary and Texts	Cologne: Rüdiger Köppe	2008
28	K. Legere & S.Msumi	Hadithi za zamani za Kikwere	TUKI, UDSM	2008
29	K. Legere & P. Mkwan'hembo	Hadithi za Kividunda 2	Ndanda Mission Press	2008

Work in Progress

- | | | |
|---|----------------------------|---------------------------------|
| 1 | J. Rugemalira & B. Phaniel | A Grammar of Mashami |
| 2 | H. Muzale | Tanzanian Sign Language Grammar |
| 3 | J. Rugemalira | Cigogo Dictionary |

APPENDIX 2: A sample of the lexical list equivalents in English (1820 out of 5000 words)

abandon	abdomen	ability
abnormal	abound	abstain
abundant	abuse	accept
accessible	accident	accomplish
accuse	acidity	acquit
actions	actor	actual
adder	adept	adorn
adornment	adult	adulterous
adze	aeroplane	affluence
after	afterbirth	afternoon
afterwards	again	age-group
agitator	agree	agreement
agriculture	aim	air
albino	all	ally
alone	alternate	amaze
amusement	angel	anger
ankle	annoy	anoint
answer	antelope	anthill
anthrax	anus	anvil
any	apologize	apparition
appear	appease	appendix
apportion	apprentice	approach
archer	area	arena
armpit	arms	army
arrange	arrive	arrow
ascend	ascension	ashes
ashore	assembly	asthma
astonishment	astrologer	attack
auction	auctioneer	average
avocet	avoid	axe

babble	baboon	backbiting
backbone	bad	badness
bag	bail	bait
bald	baldness	ball
bamboo	banjo	baptize
bar	barber	barbet
bark	barrel	basket
bat	bay	beach
bead	beads	beak
beam	bean	beans
beautiful	beauty	beckon
bed	bedbug	bee
beef	beehive	bee-sting
bee-wax	beg	beginning
beguile	behind	believe
bell	bellows	beloved
belt	bend	beneficiary
bereavement	beside	betray
betrayal	between	bewitch
bicycle	bile	bilharzias
bird	bird-lime	birth
bishop	bitch	bite
bitterness	blame	blanket
bleed	blemish	bless
blessing	blink	blister
blood	blunt	board
boast	boat	body
boil	bone	book
borrow	bottle	bow
bowl	box	boy
bracelet	brain	branch
brand	bread	break
breast	breathe	breeze

bribe	brick	bride
bridegroom	bridge	bring
broadbill	broadcast	broom
brothel	brother-in-law	bruise
bubbles	bucket	budgerigar
buffalo	bugle	building
bull	bullet	bundle
buoy	burial	burn
burn	burry	bushbuck
business	but	butcher
butterfly	buttermilk	buttock
cabbage	cactus	café
cage	calabash	calamity
call	calm	camp
canary	canine	cape
capsize	care	careful
carve	case	castrate
cat	catapult	catch
caterpillar	catfish	cattle
cattle-egret	cause	cave
ceiling	cement	centre
cereal	certainty	chameleon
chance	change	chaos
charcoal	charge	charm
chat	chatterbox	cheat
cheek	cheerful	cheese
cheetah	chest	chew
chick	chicken	chief
chiefdom	child	childishness
chill	chimpanzee	chin
cholera	church	circle
circumcise	citizen	citizenship
city	class	classification

claw	clay	clean
clear	cleverness	climb
clitoris	cloth	clothe
clothes	cloud	coal
coast	cobra	cock/rooster
cockroach	cold	colleague
colour	comb	come
comfort	commend	commerce
communication	community	compare
compensate	compensation	competition
complain	complaint	completely
compromise	consensus	connect
conquer	consciousness	consult
contentment	continue	continuity
continuous	contribution	control
conversation	cook	cooperate
cooperation	corner	cost
cotton	cough	council
count	courage	courting
courtyard	cousin	cover
cow	coward	cowardice
cow-bell	crab	crack
crafts	crawl	cream
create	cricket	crocodile
cross	crouch	crow
crowd	crown	crucial
crust	culminate	culprit
cultivate	cup	cure
curlew	curse	cursed
curve	custom	cut
dagger	dam	damage
dance	dandruff	dare
darkness	dark-red	darling

daub	daughter-in-law	dawn
day	daytime	death
debt	debtor	decay
deceive	declaration	decorate
decoration	decrease	deduct
defeat	defecate	deflower
delay	delicious	deluge
democracy	demolish	demonstrate
den	deny	descend
desist	despair	destitute
destruction	devil	dew
dewlap	dhow	diabetes
diamond	diaphragm	die
dig	dip	diplomacy
direct	direction	dirt
disagree	disappear	disappoint
discipline	discontent	discourage
disease/illness	dishonest	disinter
disobedience	disorder	disorganise
disown	disparage	displease
dispute/argue	dissuade	distance
distinguish	distribute	dive
divide	divination	divorce
doctor	doctor	dog
donkey	door	dormice
double	doubt	down
dowry	doze	drag
dragonfly	dream	dregs
dress	drink	drip
drizzle	drone	drop
drought	drum	drummer
drunk	drunkard	dry

duck	dump	duration
dust	dwelling	eagle
ear	earth	earthquake
eat	ebony-tree	echo
eel	effort	egg
egg-plant	eight	eighteen
eighty	ejaculate	eland
elbow	elder	election
elephant	eleven	eloquence
elucidate	emaciated	embark
ember	embezzle	embrace
emotion	emphasize	employ
empty	encircle	encounter
end	endurance	enemy
energy	English	enslave
enter	entrails	envelope
epiglottis	epilepsy	equality
era	erode	erosion
estimate	estrangement	eulogize
even	evening	ewe
exactly	exaggerate	example
except	exchange	excitement
excrement	expect	expensive
expert	explain	extinguish
extravagant	eye	eyebrows
eyelashes	fable	fabricate
face	fail	faint
faith	falcon	fall
fame	famine	fan
far	fart	fasting
father	father-in-law	fatigue
favour	fear	feather
feathers	feed	feel
female	fence	ferment

fever	fiancé	fierce
fierceness	fifteen	fifty
fight	fin	find
fine	finger	fire
first	fish	fisherman
fishing	fish-net	fish-trap
fist	fit	five
flag	flame	flame-tree
flatten	flea	float
floods	floor	flour
flow	flute	fly
flycatcher	fly-whisk	foam
fold	follow	food
fool	footprint	forearm
forecast	forefinger	forehead
foreman	forest	foretell
forever	forge	forget
forgetful	forgive	forgiveness
forty	forward	foundation
four	fourteen	fracture
francolin	free	Friday
friend	friendship	frighten
frog	from	frontier
fruit	fuel	funnel
fur	furnace	furrow
gall-bladder	game	gargle
garment	gazelle	generation
generosity	generous	ghee
gills	giraffe	girl
give	gizzard	glory
glue	glutton	go
goal	goat	god
gonorrhoea	good	goose
government	grab	grace

grain	granary	grandfather
grandmother	grandson	grass
grasshopper	grave	gravel
graveyard	greed	green
greet	grind	groin
groundnut	grow	growl
grumble	grunt	guava
gulf	gulp	gum
gun	gunpowder	haemorrhoids
hair	hair-dresser	half
hammer	hammer	handbag
handkerchief	hang	happen
happiness	harbour	harpoon
harrier	harvest	haste
hatch	hatred	have
hawk	head	headman
head-pad	headscarf	heal
health	hear	heart
heartburn	heat	heaven
heavy	hedgehog	heed
heel	he-goat	heifer
height	heir	helper
hem	hen	herb
herdsman	here	hernia
hero	heron	hey!
hiccup	hide	high
hike	hill	hinterland
hip	hippopotamus	hoe
hog	hole	homestead
honey	honey-badger	honeymoon
hoof	hope	hornbill
hornet	host	house
humour	hundred	hunger
hunt	hunter	hunting

hurry	hurt	husband
hut	hyena	hypocrite
hyrax	identification	ideology
idiot	ignorance	ignore
illustration	imbalance	imitation
impartial	impatient	implore
important	impossible	impotent
imprison	improve	inbreed
incapable	incentive	incest
incinerate	incise	income
increase	incurable	indebted
indeed	indicate	indigenous
inedible	ineffective	infant
infect	inform	information
ingenious	inhabitant	inherit
inheritance	injure	injustice
inquire	insane	insect
inseparable	insist	insomnia
intelligence	intelligent	intermarry
interrupt	intervene	intestine
intoxicate	intrude	invite
iron	irrigate	island
jackal	jack-fruit	jaw
jealousy	jest	jigger
job	joint	joke
journey	judge	judgement
jump	junior	kettle
key	kick	kidney
kinship	kiss	kitchen
kite	kitten	knee
kneel	knife	knot
know	knowledge	kudu
label	labour	ladder
lake	lamb	lamp

land	language	lanner
lapwing	lark	larvae
last-born	later	laugh
laughter	law	lawlessness
lawsuit	laziness	lazy
lead	leadership	leap
learn	leech	left-hand
legitimacy	lemon	lend
length	lentil	leopard
leper	leprosy	levy
liar	license	lick
lid	lie	life
light	like	lime
limit	limp	lion
lip	listen	live
liver	lizard	load
lock	locust	loft
logic	loincloth	long
loosen	lose	louse
love	lovebird	luggage
lukewarm	lump	lung
lungfish	luxury	machine
madness	maggot	magic
maize	malaria	malt
man	manage	mango
mantis	manure	market
mason	massage	mat
match	matrimony	meal
meaning	means	measles
measure	meat	medicine
meet	meeting	melt
memory	mend	menstruate
menstruation	middle	midnight
midwife	migrate	migration

milk	millet	millipede
mine	mint	miracle
mirror	miscarriage	mishap
mix	modest	mole
molest	Monday	money
mongoose	month	moon
moonlight	moorhen	moreover
morning	mosque	mosquito
mother	mother-in-law	mountain
mourning	mouse	moustache
mouth	move	Mrs
mud	mumps	murder
murmur	muscle	mushroom
musk	mute	myth
nail	nakedness	name
namesake	narrow	nausea
navel	necessity	neck
needle	negotiate	neighbour
nephew	nest	net
never	nevertheless	new
niece	night	nightjar
nightmare	nine	nineteen
ninety	no	noise
nonsense	noon	nose
nosebleed	nostril	now
numbness	nurse	oath
obey	obstruct	obtain
occupation	ochre	offer
offspring	old	one
onion	ooze	opportunity
or	orange	order
ornaments	orphan	osprey
ostrich	otter	our(s)
overcome	overnight	overpower

owe	owl	owner
oyster-catcher	pacify	pack
packet	paddle	padre
palate	palm-tree	panache
pancreas	pangolin	papaw
paper	parcel	parent
parliament	part	pass
passenger	pastor	patch
payment	pea	peace
peel	peg	pelican
penetrate	peninsular	penis
people	pepper	permit
persevere	person	pestle
phlegm	pickle	picture
piece	pierce	pig
pigeon-pea	pigmy	pillage
pillow	pillow-case	pilot
pinch	pineapple	pintail
pitfall	placenta	plan
plant	plate	play
please/satisfy	plover	plunder
pocket	poison	poke
polygamy	ponder	porcupine
potato	pottery	pour
poverty	power	praise
pray	pregnancy	prepare
pretend	previous	price
pride	prison	privacy
prohibit	promise	pronounce
property	prophet	prosper
prosperity	prostitute	protect
puberty	pull	punishment
pupil	purchase	pure
pus	push	python

quality	question	quick
quietness	quiver	rack
radish	rain	rainbow
ram	ramble	ramshackle
randy	rank	rape
rare	rat	rave
raw	read	real
rebirth	receive	recognize
reconcile	recur	red
redshank	reed	reedbuck
reeds	refuse	regrets
rehabilitate	reign	reject
relative	relish	remainder
remind	rent	repeat
repent	repetition	replace
represent	representative	repulse
resemble	reserve	resign
respond	rest	resurrect
resurrection	retaliate	reunite
revive	revolve	rhinoceros
rib	riddle	ride
ridge	right-hand	rigid
ring	ringworm	rinse
ripe	ripple	rise
rival	river	road
rob	rock	roll
roller	roof	room
root	rot	round
rub	ruin	rumble
ruminant	run	rust
sacrifice	sadness	sail
salary	saloon	salt
salvation	sand	sandal
sandpiper	sap	satiate

satisfy	Saturday	saw
say	scabies	scar
scatter	school	scissors
scold	scoop	scorpion
scram	scramble	scraper
scraps	scratch	scrub
scruffy	scrutinise	sculptor
seafaring	season	seasoning
seaweed	secede	secret
section	seduce	see
seeker	seem	seldom
selection	sell	seller
semen	semi	send
sensitive	sentry	separation
servant	service	sesame
session	set	seven
seventeen	seventy	severe
severity	sew	sewage
sex	shade	shadow
shame	shame-faced	shape
share	shatter	sheath
sheen	sheep	shell
shield	shilling	shin
shin-bone	ship	shiver
shoe	shopkeeper	shoplift
shoplifter	shortage	shortcut
shortness	short-tempered	shoulder
shovel	show	shower
shy	sickly	side
side-dish	sidestep	sieve
sift	silence	similar
simmer	simpleton	sin
since	sing	single
sinner	sip	sisal

sit	six	sixteen
sixty	size	skeleton
skill	skim	skip
skipper	skull	sky
slack	slander	slap
slaughterhouse	sleek	sleep
slip	slipperiness	slope
slowly	smallness	smear
smell	smith	smoke
snail	sneeze	sniffle
snore	snuff	soap
sob	sodomise	soldier
sole	someone	sometimes
song	son-in-law	soot
sound	sour	source
south	space	spark
sparrow	spawn	speaker
spear	spectator	speech
spider	spiel	spill
spinach	spirit	spit
spleen	splinter	sponge
spool	spoon	spouse
sprain	sprinkle	sprout
squander	squash	squat
squeeze	stagger	stair
stalk	star	starling
startle	starve	steal
steam	steenbok	steer
stepfather	stepmother	stick
sting	stinginess	stink
stir	stone	stool
stop	store	stork
story	strain	stranger
strangle	strap	stream

strength	stretcher	strict
stroke	stroke/pat	struggle
stubborn	stump	stupidity
succeed	succulent	suck
suckle	sue	sugar
sun	sunbathe	sunbeam
sunburn	Sunday	sunflower
sunrise	sunset	superstition
supervise	support	surpass
surrender	surround	suspect
swallow	swear	sweep
sweeper	sweet	swell
swelling	swift	swim
sword	syphilis	table
tablet	tail	tailor
talent	talon	tapeworm
task	taste	tattoo
taxi	TB	tea
teach	tear	tears
teenage	teeth	telephone
temple	temptation	ten
tenant	tendon	terror
terrorist	testicle	tetanus
than	thank	that
theft	their	themselves
there	thief	thigh
think	thirst	thirsty
thirty	this	thorn
thorough	thoughtfulness	thoughts
thousand	thread	three
threshold	thrice	throat
throb	thousand	throw
thrush	thud	thumb
thunder	Thursday	tick

ticket	tickle	tighten
tilapia	time	tin
tiny	tobacco	today
toe	together	tomato
tomorrow	tongue	tooth
toothbrush	top	torch
tortoise	torture	touch
tough	tow	town
trading	tradition	transparent
transplant	trap	travel
traveller	treasure	tree
trough	trousers	truth
Tuesday	tumble	turnstone
twelve	twenty	twice
twins	twist	two
type	udder	ululate
ululation	umbrella	unabashed
unable	unarmed	unbend
unbind	unbounded	unbreakable
uncap	unceasing	unchain
unconverted	uncooked	uncover
uncultivated	undeniable	undependable
underskirt	understand	underwear
undress	unemployed	unending
unengaged	unequal	unfold
ungrateful	unhappy	unhelpful
unhook	unimportant	unintelligent
unity	unkind	unknown
unload	unlock	unlucky
unmarked	unnecessary	unpack
unpopular	unprocurable	unprofitable
unreal	unrecoverable	unreliable
unscrupulous	unspeakable	until

untrue	unyielding	upbringing
uplands	upright	uproot
urinate	urine	use
uterus	uvula	vagabond
vagina	valley	vapour
vein	vendetta	vengeance
venue	vernacular	very
victim	victory	vigour
village	vindictive	virgin
virtue	visible	vision
visit	voice	vomit
wage	waist	wait
walkway	wall	want
warm	warmth	warn
warthog	wartime	wash
wasp	waste	water
waterbuck	waterproof	wave
waves	wax	wealth
wean	wear	weasel
weaver	Wednesday	weed
week	weight	well
west	wet	whale
what	when	where
whetstone	whip	whirlwind
whisper	whistle	whistling
white	whiteness	who?
wicked	wide	widow-bird
wife	wild-cat	wind
window	wing	winner
winnow	wipe	wire
wisdom	witchcraft	wither
woman	womb	wonder
wonders	woodpecker	word
work	workshop	world

worm

worship

write

yard

yawn

yaws

year

yes

yesterday

yoghourt

yoke

young

zebra

zombies

REFERENCES

- BATIBO, HERMAN. ca. 1990. MS. Questionnaire on morphological characteristics in Bantu languages. Kidadisi cha maumbo-miundo katika lugha za Kibantu. University of Dar es Salaam.
- BIBLE SOCIETY OF TANZANIA. 1960. *Shandiwka Shela sha Mulungu. The Bible in Sukuma*. Dodoma.
- BIBLE SOCIETY OF TANZANIA. 2000a. *Kyaasa Kiiya. The New Testament in Kimashami*. Dodoma.
- BIBLE SOCIETY OF TANZANIA. 2000b. *Ilagano Ipya. The New Testament in Sukuma*. Dodoma.
- BROOK-UTNE, BRIGIT, ZUBEIDA DESAI, and MARTHA QORRO, eds. 2004. *Researching the language of instruction in Tanzania and South Africa*. Cape Town: African Minds.
- CASTON, BRIAN, SHERI DAGGETT, TRACY HADLOCK, and KATHERIN TURNER. 1996. *Sociolinguistic survey among the Isanzu people, Iramba District, Singida Region, Tanzania*. Dodoma: Summer Institute of Linguistics.
- CASTON, BRIAN, SHERI DAGGETT, TRACY HADLOCK, and KATHERIN TURNER. 1997. *Sociolinguistic survey among the Daiso/Segeju, Muheza District, Tanga Region, Tanzania*. Dodoma: Summer Institute of Linguistics.
- EATON, HELEN. 2005. Focus as a key to the grammar of Sandawe. *Occasional Papers in Linguistics* 1. Dar es Salaam: LOT Project, University of Dar es Salaam.
- EGERÖ, BERTIL, and ROUSHDI A. HENIN, eds. 1971. The population of Tanzania: An analysis of the 1967 population census. *Census Volume 6*. Dar es Salaam: BRALUP and Bureau of Statistics.
- ENDL, MICHAEL, and LIZ THOMSON. 2002. *Giyaandimiisay gu handikiru gu Burungaisoo. Mwongozo wa Mwandishi wa Kiburunge (Burunge writer's guide)*. Dodoma: SIL International.
- FACULTY OF EDUCATION. 2006. A special edition on language of instruction. *Papers in Education and Development* 26. University of Dar es Salaam.
- GADELI, KARL. 2001. Languages in Mozambique. *Africa & Asia: Göteborg Working Papers on Asian and African Languages and Literatures* 2:163–186. Göteborg University.
- GRIMES, BARBARA, ed. 2000. *Ethnologue: Languages of the world*. 14th Edition. Dallas: SIL International.
- GUTHRIE, MALCOLM. 1948. *The classification of the Bantu languages*. London: Oxford University Press.
- GUTHRIE, MALCOLM. 1967–1971. *Comparative Bantu*, volumes 1–4. Farnborough, England: Gregg International Publishers.
- KAGAYA, RYOHEI. 1993. *A classified vocabulary of the Sandawe language*. Tokyo: ILCAA.
- LANGUAGES OF TANZANIA PROJECT. 2005. *Hojaji ya Isimu-Jamii*. MS. Dar es Salaam: LOT Project, University of Dar es Salaam.
- LEGERE, KARSTEN. 2002. The “Languages of Tanzania” Project: Background, resources, and perspectives. *Africa & Asia: Göteborg Working Papers on Asian and African Languages and Literatures* 2:6–12. Göteborg University.
- MAHO, JOUNI, and BONNY SANDS. 2002. *A linguistic bibliography of Tanzania*. Göteborg: Göteborg University.

- MASSAMBA, DAVID. 2005. *Eciruuri lexicon*. LOT Publications Lexicon Series No. 3. Dar es Salaam: LOT Project, University of Dar es Salaam.
- MOUS, MAARTEN, MARTHA QORRO, and ROLAND KIESLING. 2002. *Iraqw-English dictionary*. Köln: Rudiger Koppe Verlag.
- MSANJILA, YOHANA. 1999. *The use of Kiswahili in rural areas and its implications for the future of ethnic languages in Tanzania*. PhD diss., Department of Kiswahili, University of Dar es Salaam.
- MEKACHA, RUGATIRI. 1993. *The sociolinguistic impact of Kiswahili on ethnic community languages in Tanzania: A case study of Ekinata*. Bayreuth: University of Bayreuth.
- MKUDE, DANIEL. 2002. Minority languages and democratization in the SADC region: The case of Tanzania. In *Talking freedom*, ed. by Karsten Legere and Sandra Fitchat, 67–76. Windhoek: Gamsberg Publishers.
- MOE, RONALD. 2002. Lexicography and mass production. Paper presented at the Languages of Tanzania Project Workshop, 8–9 March 2002. University of Dar es Salaam.
- MRETA, ABEL, HENRY MUZALE and JOSEPHAT RUGEMALIRA. 2002. *Languages of Tanzania Project: Workshop II Report*. University of Dar es Salaam.
- MUZALE, HENRY. 2006. *Oruhaya dictionary*. LOT Publications Lexicon Series No. 5. Dar es Salaam: LOT Project, University of Dar es Salaam.
- MUZALE, HENRY, and JOSEPHAT RUGEMALIRA. 2001. Languages of Tanzania Project: Workshop I Final Report. University of Dar es Salaam.
- Papers in Education and Development*. 2006. Journal of the Faculty of Education 26. University of Dar es Salaam.
- ROY-CAMPBELL, ZALINE and MARTHA QORRO. 1997. *Language crisis in Tanzania: The myth of English versus education*. Dar es Salaam: Mkuki na Nyota Publishers.
- RUBAGUMYA, CASMIR, ed. 1990. *Language in education in Africa: A Tanzanian perspective*. Clevedon: Multilingual Matters.
- RUBAGUMYA, CASMIR. 2006. *Igihangaza lexicon*. LOT Publications Lexicon Series No. 4. Dar es Salaam: LOT Project, University of Dar es Salaam.
- RUGEMALIRA, JOSEPHAT. 2002. *Orunyambo lexicon*. LOT Publications Lexicon Series No. 1. Dar es Salaam: LOT Project, University of Dar es Salaam.
- RUGEMALIRA, JOSEPHAT. 2004. Mapping the languages of Tanzania, in *Journal of the African Language Teachers' Association* 5 (1):13–19.
- RUGEMALIRA, JOSEPHAT. 2005a. *A grammar of Runyambo*. LOT Publications Grammar Series No. 1. Dar es Salaam: LOT Project, University of Dar es Salaam.
- RUGEMALIRA, JOSEPHAT. 2005b. Theoretical and practical challenges in a Tanzanian English medium primary school. *Africa & Asia: Göteborg Working Papers on Asian and African Languages and Literatures* 5, 66–84. Göteborg University.
- TANZANIA GOVERNMENT. 1997. *Sera ya Utamaduni. Cultural Policy*. Dar es Salaam.
- TANZANIA GOVERNMENT. 2003. *2002 Population and housing census: General report*. Dar es Salaam: Government Printer.
- TURNER, KATHERIN, DAVID LIDDLE, ANNEMARIE BURKE, SHERI DAGGETT, and TRACY HADLOCK. 1998. *Sociolinguistic survey among the Safwa people, Mbeya District, Mbeya Region, Tanzania*. Dodoma: Summer Institute of Linguistics.