4

Linguistic variation and the dynamics of language documentation: Editing in 'pure' Kagulu[1]

Lutz Marten *(SOAS, University of London)*
and Malin Petzell *(University of Gothenburg)*

The Tanzanian ethnic community language Kagulu is in extended language contact with the national language Swahili and other neighbouring community languages. The effects of contact are seen in vocabulary and structure, leading to a high degree of linguistic variation and to the development of distinct varieties of 'pure' and 'mixed' Kagulu. A comprehensive documentation of the language needs to take this variation into account and to provide a description of the different varieties and their interaction. The paper illustrates this point by charting the development of a specific text within a language documentation project. A comparison of three versions of the text – a recorded oral story, a transcribed version of it and a further, edited version in which features of pure Kagulu are edited in – shows the dynamics of how the different versions of the text interact and provides a detailed picture of linguistic variation and of speakers' use and exploitation of it. We show that all versions of the text are valid, 'authentic' representations of their own linguistic reality, and how all three of them, and the processes of their genesis, are an integral part of a comprehensive documentation of Kagulu and its linguistic ecology.

1. INTRODUCTION. Language contact is an important aspect of and a precondition for language shift and language endangerment, therefore language documentation must take heed of language contact, its effects on language structure, and the sociolinguistic spaces that language contact provides (cf. Childs et al. 2013). More generally, language documentation aims to provide a comprehensive account of variation encountered in the language, and of different linguistic forms associated with different contexts of language use (e.g. Himmelmann 1998, 2006).

Language contact situations and linguistic variation exist independently of language description, documentation or revitalization efforts. However, linguistic variation can acquire a particular dynamic in language documentation contexts. In this paper we will present a detailed example of this from the multilingual context of Tanzania, showing the emergence and interaction of different language forms in different situations. The discussion is based on a specific example from the Tanzanian ethnic community language Kagulu (G12, 336,749 speakers).[2] It charts the development and editorial processes in the

---

[2] The classification of the languages discussed in the article follows Maho's (2003) update of Guthrie (1967-71); speaker numbers are from MLT (2009).

genesis of a particular Kagulu text in the context of a language documentation project. By comparing three versions of the text – a recording of an oral version, a transcribed version of the recording, and an edited version of the transcription – we show the dynamics of language variation and speaker agency and trace the construction of a 'pure' Kagulu text (cf. Mosel 2008, 2012, Woodbury 2005, 2011). The example shows how through the agency of members of the documentation team, different language use and language ideologies take effect and the nature of the text is transformed from the initial recording of an oral story into a written and edited version of the text. In the process, the text is 'purified' in that perceived effects of language contact with Swahili are replaced by forms seen as being more 'authentic' Kagulu. In addition, variant forms are standardized, and the genre of the text as a story is reinforced through the introduction of genre-specific forms.

We will propose that all versions of the story are representations of authentic Kagulu and that it is precisely the dynamics of the development of the different versions, and the interaction of language contact, standardization, spoken vs. written text creations, and speaker agency (particularly, in this case, within a language documentation project) which provides the context in which to interpret and understand the particular language ecology and language dynamics of Kagulu. Both the texts themselves and the contexts of their production thus provide essential evidence for language documentation.

2. BACKGROUND. Tanzania is home to a complex linguistic situation, involving between 120 and 150 languages belonging to four different linguistic phyla (e.g. Kahigi et al. 2000, Lewis et al. 2014, Maho & Sands 2002, MLT 2009). The largest language, in terms of first language speakers, is Sukuma, with just over five million speakers, but speaker numbers of most languages are in the hundred thousands rather than millions.

The most dominant language of the Tanzanian linguistic situation is Swahili. Originally spoken along the East African coast from southern Somalia to northern Mozambique, in modern times Swahili has become the main East African lingua franca, with high numbers of speakers in Kenya, Tanzania, Uganda, parts of the Democratic Republic of the Congo and Mozambique. The role of Swahili as a lingua franca goes back to the 19th century, and the language has been strongly promoted as the national language of Tanzania since the foundation of the republic in 1964 (Mbaabu 1985, Whiteley 1969). It is now established as the almost universal lingua franca of the country and functions as the main or only language of public discourse, the media, education, commerce and government. Through widespread use and its prevalent role in education for the last four decades, knowledge of Swahili in Tanzania, often as a second language, has become near-universal. The majority of Swahili speakers are bi- or multilingual in Swahili and in an ethnic community language, which has typically been their home or first language. This is changing, however, especially in urban contexts, where Swahili has often become a first language of younger speakers.

Tanzania's ethnic community languages are largely marginalized and sometimes actively suppressed (Muzale & Rugemalira 2008). They have no designated roles in public life and are mostly confined to home use. They are not used in education, in the media or in other public domains, and their use in the classroom or at political rallies is discouraged. Few have an agreed orthography or a written corpus, and only recently have there been efforts to map and document ethnic community languages (e.g. Legère 2002, MLT 2009).

The overall language situation is characterized by a high degree of multilingualism, typi-

cally involving Swahili and one or more community languages, and by language shift and loss, typically away from an ethnic community language to Swahili (Batibo 1992, Kiango 2013, Legère 1992, Mekacha 1993, Petzell 2012a, Yoneda 1996). Of the 126 Tanzanian languages listed in Lewis et al. (2014), 40 are classified as being 'in trouble', and eight as 'dying', and there are many documented cases of the effects of language contact and the influence of Swahili on ethnic community languages. As in many other African (and non-African) contexts, gradual language shift, rather than abrupt language loss or extinction, is the main process of language endangerment in Tanzania, and language shift typically involves an African lingua franca, in this case Swahili, rather than an ex-colonial language (e.g. Batibo 2005). In addition, and sometimes in parallel with language shift, many Tanzanian community languages show signs of language contact with Swahili. Aspects of the 'Swahilization' of Tanzanian community languages are found at all levels of linguistic structure – vocabulary, phonology, morphology and syntax. Often there is a difference in usage between generations, with older speakers speaking a more conservative variety.[3] Speakers are often aware of differences between what are seen as 'pure' or 'deep' versus 'modern' or 'mixed' varieties (e.g. Yoneda 2010; see also Slabbert & Finlayson 1998 for this distinction in a South African context).[4]

For example, Yoneda (2010) reports that contact effects in the south-western Tanzanian language Matengo (N13, about 270,000 speakers) are found in phonology, vocabulary and morphosyntax. Although Matengo has a seven-vowel system with a distinction between short and long vowels, younger speakers are shifting towards a five-vowel system without length distinctions, which corresponds to the vowel system of Swahili. In the lexicon, many loanwords are taken from Swahili, not only for new concepts, but also for meanings for which a synonymous Matengo word exists. Loanwords are not restricted to open class items, but also include function words such as prepositions, conjunctions and adverbs. On the grammatical level, Matengo has borrowed a passive construction from Swahili, replacing an impersonal construction (in effect, this change is a re-introduction of a historical construction which had been lost, since the Swahili passive marker is a reflex of a reconstructed Proto-Bantu form). Matengo speakers are well aware of these on-going changes and distinguish between *Samatengo sa ndani* ('pure' or 'inside' Matengo) which is considered as 'real' Matengo, and *Samatengo sa kisasa* ('modern' Matengo), which shows strong influence from Swahili (Yoneda 2010: 147).

In Digo (E73, 166,000 speakers), spoken in the coastal region along the Kenyan-Tanzanian border, Nicolle (2013: 413–417) notes influence from Swahili in vocabulary and grammar. There are several loanwords from Swahili, some of which have become part of the 'core' vocabulary, and are often used alongside older Digo words. Grammatical influence from Swahili can be seen in the use of demonstratives, where only older speakers make full use of the Digo system of four series of demonstratives (with different variant forms), a situation reflecting more closely the Proto-Bantu system as de-

---

[3] Or at least the difference is often reported as being related to age. Education, profession or place of residence (e.g. urban vs. rural) are likely to play a role as well, and there is an additional dimension of observed vs. actual language use. We will see in the Kagulu example below that the promotion of more conservative language forms is actually found among younger speakers.

[4] There are also inverse contact effects, where features from ethnic community languages (e.g. diminutive forms with class 12 *ka-*, habitual *-ag-*) are incorporated into colloquial varieties of Swahili (Marten 2013, Rugemalira 2010).

scribed by Meeussen (1967: 107), while younger speakers use fewer forms, resembling the three-way distinction of Swahili. Another area is in the use of noun class 11, which is changing in terms of variation of forms and plural assignment, possibly under influence from Swahili, where classes 11 and 14 have merged. However, while there are contact effects, Nicolle observes that there is no evidence of wholesale language shift from Digo to Swahili, even though the domains in which Digo is used are limited.

The neighbouring language Bondei (G24, 121,934 speakers) has likewise been in contact with Swahili for a long time, and shows signs of language contact in all areas of language structure, and particularly in the lexicon. Kiango (2013) notes that Bondei is used as the main means of communication only in a small number of villages, and typically by older speakers. In contrast, in most villages and all urban contexts, both Bondei and to a larger extent Swahili are used, and most younger speakers are more likely to use Swahili than Bondei. While younger speakers typically understand Bondei, when speaking they often mix Bondei with Swahili forms.

The examples of Matengo, Digo and Bondei are in many ways representative of the wider language situation in Tanzania. Tanzanian ethnic community languages are restricted in their domains of use, often show contact effects, mainly from contact with Swahili, and experience more or less language shift and language endangerment. Furthermore, the vast majority of Tanzanian languages remains underdocumented despite recent efforts in increased linguistic description. A problem for language documentation in this context is the representation of the different varieties along the poles of 'pure' and 'modern' varieties, and how to capture adequately the dynamics underlying the relevant contact situation. On the one hand, documentation aims at presenting the language's lexical and structural resources as extensively as possible, including forms and features of 'pure' varieties or 'ancestral code' (Woodbury 2005, 2011). On the other hand, documentation has to be faithful to actual language use and linguistic practices embedded in contemporary communicative ecologies, and to the heterogeneity and dynamics inherent in any language – thus taking account of 'modern' varieties as well (Himmelmann 1998). The situation is complicated by the fact that speakers are often aware of differences between different varieties, and may consciously or unconsciously manipulate certain forms or codes rather than others on a given occasion – for example, in response to the presence of a researcher, or so as to present, or create, a more distinct and unique variety, strengthening a distinct linguistic identity, for example in the context of producing edited, written versions of spoken texts (Mosel 2008, 2012), as we will show below.

Approaches to this problem have already been seen in the short discussion above. The descriptions by Yoneda (2010), Nicolle (2013) and Kiango (2013) provide explicit statements about structural influences from Swahili on the structure of Matengo, Digo and Bondei, respectively. In addition, the authors comment on differences in use, in particular with respect to different generations, and on speakers' attitudes towards the different varieties. In Yoneda's case, this is backed up by interviews conducted with groups of younger and older speakers (2010: 146–7). The approach adopted by these authors is to provide explicit documentation of the variation and dynamics found in the contact situation, akin to the aims of 'sociolinguistic language documentation' developed by Childs et al. (2013).

In the following section, we consider another means of capturing the dynamics of language contact in a language documentation context, and that is the genesis of different variants of a Kagulu text – an original recording, a transcribed version, and an edited text – as

a record of editing in 'pure' Kagulu.

3. *DIFISI NA SUNGULA* 'THE HYENA AND THE RABBIT': THE GENESIS OF A KAGULU TEXT. In this section we present a detailed analysis of the genesis of an edited Kagulu text within a language documentation project. We will provide a brief background of the sociolinguistic situation of Kagulu and then present and compare the three versions of the text – the recording, the transcription and the edited text – in detail, showing how speaker agency, language attitudes and changes in mode from spoken to written result in the dynamic development of the text.

3.1 KAGULU LANGUAGE BACKGROUND. Kagulu (autonyms *Chikagulu* or *Chimegi*) is a Bantu language spoken in Tanzania's Morogoro region, about 250 km from the coast and the country's largest city Dar es Salaam, with some 336,749 speakers (MLT 2009; cf. Petzell 2008, 2012b). There are a number of anthropological works on the language, including Kagulu language materials (e.g. Beidelman 1967, 1971, 1997), two grammars (Last 1886, Petzell 2008) and a recent corpus resulting from documentation work by Malin Petzell, part of which is included in the Kagulu deposit at the Endangered Languages Archive (ELAR) at SOAS (http://elar.soas.ac.uk/deposit/petzell2010kagulu). Our entire Kagulu corpus consists of approximately 15 transcribed stories and conversations, numerous in-depth interviews and elicitation sessions, and months of participant observation in the field over a period of more than ten years. Our discussion here is based on this corpus.

Consonant with the foregoing discussion of ethnic community languages, Kagulu is restricted to a few domains of use, and its use is actively discouraged in schools and other public functions. There are high levels of Kagulu-Swahili bilingualism, and Swahili is used as the language of education, the media and politics. However, the language is being transmitted to children, and speakers have overall positive attitudes towards the language. The Kagulu-speaking area is surrounded by several other languages. Zigula (G31), Ngh'wele (G32), Zaramo (G33), Ngulu (G34), Luguru (G35) and Sagala (G39) are Bantu languages spoken to the east and southeast, while Gogo (G11) and Hehe (G62) are spoken to the west. The only neighbouring non-Bantu language is the Nilotic language Maasai, spoken to the north. Kagulu speakers are often conversant with neighbouring languages, some of which are mutually intelligible with Kagulu, and there is some contact-induced influence on Kagulu from neighbouring languages (Petzell 2008: 25–6).

As in other Tanzanian community languages, there are several signs of language contact with Swahili, and linguistic influence can be seen at the lexical and grammatical level. Like Matengo speakers, noted above, Kagulu speakers distinguish between 'pure' Kagulu (*Chimegi muhala* 'only Kagulu' or *Chimegi chenyecho* 'Kagulu itself') and 'modern' varieties, which are referred to as 'mixed language' (i.e. a mix with Swahili).

There are numerous Swahili loanwords, including lexical borrowing of nouns, verbs and adjectives (Petzell 2008: 80), as well as numerals and borrowed function words such as the complementizers *kwamba* and *yani* 'that' (2008: 187), and the question word *nani* 'who', which, in contrast to its Kagulu counterparts, is uninflected (2008: 177). In the noun class system, the assignment of class 14 plurals to class 10 is likely to be an innovation, and influenced by Swahili (2008: 62), while the reduced use of the augment (a vocalic 'pre-prefix' (PP) of the noun class prefix) by younger and urban speakers may reflect the Swahili sys-

tem, which does not have augments (2008: 67). In verbal morphology, there are alternations in the shape of the causative (*-is-* vs. *-ish-*) and passive (*-igw-* vs. *-w-*) markers, which are probably due to the introduction of the Swahili forms *-ish-* and *-w-* (2008: 119, 135). In comparative constructions, most speakers use *kuliko* 'than' (1a) in daily speech – a loan from Swahili – while the dated form *kusuma* 'exceed' (1b) (2008: 81) had to be elicited:[5,6]

(1) a. i-biki   di-no   i-tali   kuliko di-monga
         5-tree   5-DEM   5-tall   than   5-other
         'This tree (is) taller than the other one'

     b. m-gosi   m-kulu   ku-sum-a   i-mu-jelelo
         1-man   1-big   15-exceed-FV   PP-1-young.man
         'The man (is) bigger than the boy'

In copula constructions, there is an alternation between an older copula form *no*, which is associated predominantly with use by rural and/or older speakers and is found in historic (19th century) texts as well as modern ones, and a newer form *ni*, used by all speakers, which is not attested in older sources. Since it is the same form as the copula in Swahili, the *ni* form is probably due to contact. Despite a slight tendency for rural and/or older speakers to use the older form *no*, all in all the two copulas appear to be in free variation (2008: 162), and can even be used by the same speaker in the same phrase:

(2) kasi   s-akwe   no   si-s-o   ni   ng'hefu
     10.work   10-3SG.POSS   COP   10-10-REL   COP   10.few
     'His/her works are those which are few'

The examples illustrate some of the structural influence Swahili has on Kagulu. In many cases, variation between two different forms can be associated with the speaker's level of exposure to Swahili, which is usually higher among younger, urban, more educated speakers.

Kagulu speakers are often aware that they mix Kagulu with Swahili (and occasionally English) and may correct others as well as themselves. In the conversation between two speakers below, the speakers comment on their use of English and Swahili words and start correcting each other (laughingly) as they speak:

(3) a. Goloko u-ku-changany-a.
         friend   2SG-PRES-mix-FV
         'My friend, you are mixing.'

---

[5] Glossing conventions follow the Leipzig Glossing Rules. In addition, the following abbreviations are used: 1, 2, 3 = noun class number, ACP = agreement class prefix, CONJ = conjunction, EXT = extension, FV = final vowel, INT = interrogative, INTERJ = interjection, PP = pre-prefix, N = noun, NCP = nominal class prefix, NUM = numeral, OM = object marker, PoS = part of speech, PRN = pronoun, REF = referential, S = sentence, SM = subject marker, TM = tense marker, V = verb.

[6] The adjective *-tali* 'tall' in (1a) looks like a loan from English, but this is a chance resemblance and it can be reconstructed as Proto-Bantu **tadi*.

    b.   Hambiya   lugha   si-etu   si-no   si-o-baho
           now   10.language   10-2PL.POSS   10-DEM   10-REL-DEM

           si-ki-ingil-il-a.
           10-REC-enter-APPL-FV
           'Well, it is here our languages interconnect.'

In (3), *-changanya* 'mix' and *lugha* 'language' are loanwords from Swahili, the corresponding Kagulu forms being *-hasa* 'mix' and *nonga* 'language'.

3.2 *DIFISI NA SUNGULA:* ALL FIVE VERSIONS OF THE TEXT. The interplay between the different forms of Kagulu – 'pure' Kagulu, associated with less influence from Swahili, and 'mixed' Kagulu, associated with stronger influence from Swahili – is particularly clear in the study of textual editing to which we now turn. The text we use to illustrate our point is a story called *Difisi na sungula* 'The hyena and the rabbit', which was recorded on 28 August 2003 by Malin Petzell in Tanzania. A sound file of the recording, as well as a transcribed (and edited) version of the story are available online as part of the ELAR Kagulu deposit, and a complete transcription, highlighting editorial changes, is provided in the Appendix. The text is broadly representative of our Kagulu data (cf. Petzell 2008, 2015).

    The story was originally told by the narrator JM and recorded in Berega, a village in the centre of the Kagulu area. It was then transcribed by a Kagulu speaker (RM) in Dar es Salaam, and typed there by a typist. The transcriber RM made some comments on the text and suggested some corrections and additions, which we will discuss in more detail below. An edited version was then produced by a different Kagulu speaker (SL) in Morogoro, working from the typed version without access to the original recording. During the editing processes, a number of changes, replacements and additions were made. Finally a new transcription of the original recording was made in 2013 by the authors of the present paper in conjunction with Nancy C. Kula (LM, MP, NCK). As shown in Table 1, there are thus five versions of the text, although we will mainly be concerned with the original text (as represented by our own new transcription), the handwritten transcribed text, and the edited text.

Table 1. Summary of the five versions of the text

Text No	Text Name	Originator	Date	Comments
1	Original recording	JM	28 August 2003	The narrator JM is female, was born in 1943 in Berega village and was living there at the time of the recording. The recordings took place in the researcher's rented house in the village.

2	Handwritten transcription	RM	September 2003	The transcriber RM is male and was born in 1969 in Berega village, but has lived and worked in Dar es Salaam. He was studying in the MA Linguistics programme at the University of Dar es Salaam at the time. RM added a number of annotations to the original text.
3	Typed transcription		September 2003	Other than the transferral from handwritten to typed format, this is identical to Text 2.
4	Edited version	SL	2004-2005	The editor is SL, born in 1970 in Chanjale near Nongwe (deep in the Kagulu area). He was living and studying in Morogoro town at the time. He was also a member of the review board assessing the Kagulu Bible translation. Numerous changes to the original text are edited in.
5	New transcription	LM, MP, NCK	19 May 2013	New transcription and the most faithful representation of the original recording (Text 1). In effect, we are treating this here as Text 1 and will use Text 1/5 to refer to this version.

3.3 FROM TEXT 1/5 TO TEXT 2. In the first transcription process from Text 1 to Text 2, the transcriber (RM) provided several annotations and suggestions for changes in vocabulary, phrasing and morphosyntax. These included the suggested standardized spelling of the Swahili form *halafu* 'then' instead of the variant forms *harafu*, *ha'afu* found in Text 1 (but in contrast to the later edited version, there is no suggestion of using the Kagulu form *kamei* 'then' instead of *halafu*), and the change from the colloquial Kagulu form *halika* 'if' to the standard form *ng'halika* 'if'. The spelling <ng'h> represents a voiceless velar nasal, which is sometimes reduced to /h/ in contemporary spoken Kagulu. RM also suggested replacing the Swahili-influenced clause *usiku uja* 'night comes' with the Kagulu adverbial *nhechilo* 'at night'. Additionally, he proposed changing the class for concord agreement for animals. In the original Text 1, agreement with animals is in class 1, otherwise mostly reserved for humans, as is the case in Swahili. In contrast, in Kagulu, agreement with the inherent noun class is used, and this is what RM suggests. Interestingly, class 1 concord was retained in the subsequently edited Text 4, possibly because the editor felt that in traditional Kagulu, nouns referring to animals may take the prefix of class 1 if the animal is personified and assigned human traits, like an animal in a fable (Petzell 2008: 48). Finally, the transcriber suggested the addition of the traditional Kagulu ending to the story, the closing formula *simo ihela* 'the story is finished' replacing the Swahili *hadithi imeisha* 'the story is finished'. These suggested changes are summarized in Table 2.

TABLE 2. Suggested changes in the transcribed version Text 2

PoS	Form in original text (Text 1/5)	Meaning	Suggested form in transcribed text (Text 2)	Notes
Adv	*halafu, harafu, ha'afu*	'(and) then'	*halafu*	Use of standard spelling, cf. also Swahili *halafu*
Conj	*halika*	'if'	*ng'halika*	Use of the standard form
S/Adv	*usiku uja*	'night come'	*nhechilo* ('at night')	cf. Swahili *usiku ukaja* ('and night came')
ACP	e.g. *ku-m-gamba* (class 1 agreement)	'to tell it (i.e. hyena)'	*ku-di-gamba* (class 5 agreement)	Agreement with nouns denoting personified animals
S	*hadithi imeisha*	'the story has ended'	*simo ihela*	cf. Swahili *hadithi imeisha*

The changes proposed by RM include forms of different parts of speech and grammatical forms, such as adverbs, conjunctions, phrases, and agreement markers, and can be grouped into three broad categories: 1) the use of standard forms for both Swahili (*halafu*) and Kagulu (*ng'ahlika*) forms, 2) the use of Kagulu rather than Swahili morphosyntax in the agreement with nouns denoting personified animals, and the replacement of the Swahili form with the Kagulu adverbial phrase *nhechilo* 'at night', and 3) the use of the Kagulu closing formula for the story instead of the Swahili one. Similar examples, and an overall trend towards standardization and purification, can be seen in the next stage of the genesis of the text, discussed in the next section.

3.4 FROM TEXT 1/5 AND TEXT 2 TO TEXT 4. During the subsequent development of the text, Text 2 was typed, resulting in Text 3, without any further changes. Text 3 then fed into a further editing process by the editor SL, who had no access to the original recording. Briefed only to proofread the text, SL added a number of changes to the text, resulting in Text 4. Overall, these new editorial changes introduced forms which seem to be more 'proper' Kagulu, i.e. *Chikagulu muhala*, 'pure' or 'only' Kagulu. The editor's revisions and corrections often consist of replacing what were probably taken to be contact-induced forms with 'pure' Kagulu forms. In the extract from the text in (4) – the first sentence and opening of the story – we have included the transcription of the original, spoken version (Text 1/5) in line 1, alongside the edited, written version (Text 4) in line 2. Differences between the two versions are highlighted in bold and grey shading. The first two words, *baho katali*, for example, are part of the edited version, but not of the original spoken version:[7]

---

[7] Lines 3 to 5 show the morphological structure of the edited version, morphological glosses and part of speech tagging. The full text of the story is provided in the Appendix.

(4) Opening of *Difisi na sungula*

		difisi	na	sungula
Baho	**katali**	difisi	na	sungula
baho	katali	di-fisi	na	sungula
DEM	long time ago	5-hyena	CONJ	9.hare
prn	*adv*	*ncp-n*	*conj*	*n*

hawowa	mbuya	**ha(l)afu**
hawowa	mbuya	**kamei**
ha-wa-uw-a	mbuya	kamei
PST-2-be-FV	1a.friend	then
tm-sm-v-fv	*n*	*adv*

sungula	kamgamba,	"chigende	**safali**."
sungula	kamgamba,	"chigende	**nhambo**."
sungula	ka-m-gamb-a	chi-gend-e	N-tambo
9.hare	1.PST-1-speak-FV	7-go-FV	9.journey
n	*sm-om-v-fv*	*sm-v-fv*	*ncp-n*

'A long time ago, the hyena and rabbit were friends, then rabbit told the hyena "let us have a journey"'.

The extract shows three differences between the two versions. As noted above, the opening formula *baho katali* 'a long time ago' was not part of the original spoken version. The second difference is the use of the adverb *kamei* 'then' in Text 4, instead of *halafu* 'then' in the original spoken version Text 1/5, the latter being a Swahili form. The third difference is the substitution of the noun *nhambo* 'journey' for *safali* 'journey', which is from Swahili *safari* 'journey'. The /r/ in the Swahili form *safari* is substituted by /l/ in *safali*, which is a frequent and regular process of adaptation, since Kagulu, in contrast to Swahili, does not use /r/ at all. The last two examples – the substitution of *halafu* and *safali* – show that the editorial process involved a process of purification, or 'de-Swahilization' and changed the original text from a more 'mixed' to a more 'pure' Kagulu. The first difference in (4), the addition of the opening formula *baho katali*, is consistent with this observation, since the addition of the opening formula adds an idiomatic, 'pure' expression to the text and adds to its 'authenticity' as a 'traditional' Kagulu story.

The observations made in the light of the first sentence hold true, by and large, for the text overall. The differences between the original recording, Text 1/5, and the edited version, Text 4, are summarized in Table 3.

TABLE 3. Changes in the edited version Text 4

PoS	Form in original text (Text 1/5)	Meaning	Form in edited text (Text 4)	Notes
Adv	*halafu, harafu, ha'afu*	'(and) then'	*kamei*	cf. Swahili *halafu*
N	*safali*	'journey'	*nhambo*	cf. Swahili *safari*
Int	*nani*	'who'	*ye-hoki*	cf. Swahili *nani*
N	*mwenye nyumba*	'house owner'	*mwinya ikaya*	cf. Swahili *mwenye nyumba*
V	*wa-kalibish-igw-a*	'they were welcomed'	*wa-hokel-igw-a*	cf. Swahili *wa-li-karibish-w-a*
V	*ya-ku-m-tamil-a*	'he tells him'	*ho-yo-m-tamil-a*	Use of past tense prefix *ha-/ho-*
S	*hadithi imeisha*	'the story has ended'	*simo ihela*	cf. Swahili *hadithi imeisha*
Adv	[none]	'A long time ago'	*Baho katali*	Addition of opening formula
Prn/ Num	*imonga*	'someone'	*imwedu/imwe*	Indefinite pronoun replaced by numeral 'one'
Conj	*kifwa*	'that'	*fina*	Replacement of conjunction

The examples show that several Swahili loanwords have been edited out. The variation in the form of the adverb *halafu*, *harafu*, *ha'afu* in the original recording shows degrees of loanword adaptation. In the edited text, all instances of the adverb are replaced by *kamei*. This has considerable influence on the text overall, as these forms are very frequent; there are 12 instances of *halafu* (and variants) in the original recording, out of 113 words in total; that is just over 10%. The replacement of *safali* by *nhambo* has been noted above, and the change of *nani* to *ye-hoki* follows a similar pattern, where the Swahili form is replaced by a more traditional Kagulu form. The form *mwenye nyumba*, from Swahili 'homeowner' (literally 'one having a/the house') is replaced by what looks like a calque translation with the same structure, *mwinya ikaya*. Other changes, with the effect of making the text more 'pure', are seen in the replacement of the (adapted) Swahili verb *kalibishigwa* (albeit with Kagulu concord) with the Kagulu *hokeligwa* (from the Kagulu verb *-hokela* 'receive'), and in *hoyomtamila*, where the verb form has been changed from present tense to past tense. The reason for this may be partly so as to use the past tense morpheme *ha-* (realized as *ho-* due to vowel assimilation), which is typical of Kagulu and is a feature not shared with neighbouring languages, including Swahili.

Other editorial changes are less frequent. One is the addition of the opening formula *baho katali*, discussed above, and of the Kagulu ending *simo ihela* instead of the Swahili

hadithi imeisha, already suggested in Text 2. Other examples include the replacement of the Kagulu indefinite pronoun *imonga* 'someone' (*dimonga* in class 5) with the numeral 'one' *imwedu* (or its short form *imwe*). This replacement does not seem to be clearly motivated, since the indefinite pronoun can be used as meaning 'one' or 'the other one'. Neither of the forms are related to the corresponding Swahili forms *-o-ote* 'anyone' or *-moja* 'one'. Similarly, in Text 4 the conjunction *kifwa* 'that' is replaced by the form *fina*. Neither form is related to a corresponding Swahili form, and the reason for the change remains unclear. It is possible that in both cases, the form in Text 1/5 was seen as inappropriate, possibly reflecting contact with a neighbouring language other than Swahili.

3.5 DISCUSSION. There are several factors to be taken into account when comparing the different versions of the text, including a change of mode from spoken to written language, the different roles adopted by the narrator, the transcriber and the editor, differences in their language attitudes, education and literacy ideologies, and the effects on the different intended audiences.

The original recording is of an oral, spoken text, while the transcription and the edited text are written. The change of mode from oral to written allows the transcriber and the editor to reflect on and change the language used, letting in the influence of different language attitudes and ideologies. A number of changes observed in the genesis of the Kagulu text are reminiscent of processes of text development, transcribing and editing observed by Mosel (2008, 2012) in the context of materials development for the teaching of endangered languages. Based on evidence from a documentation project of Teop in Papua New Guinea, and in particular of writing and editing Teop oral legends, Mosel identifies various types of changes, including purification and the replacement of loan words, elaboration through the addition of words, phrases and clauses, and increasing text coherence through different structural changes such as complexification or more explicit clause linkage. As shown above, very similar alterations are represented in the changes from Text 1/5 to Texts 2 and 4.

The change of mode from oral to written described here with reference to Kagulu is also relevant to the wider discussion of the representation of oral language, processes of transcription and the attendant transformation of the language used. Hinton (2011), for example, raises a number of issues related to writing and writing processes in language documentation and how these relate to oral language use. Ochs (1979) discusses how details of the representation of spoken language in written form reflect the cultural and theoretical assumptions of the transcriber, while Duranti (1997: 122-161) notes the relationship between writing and standardization, evident also in the case of the Kagulu texts. Geider (2003, esp. 137-146) comments on the process of text formation when writing Kanuri oral texts and identifies different stages of transition – from speech event, to recording, to the edited version of the text – sketching a similar process to the one described here. As these previous works note, writing is the most common process of reification of oral texts, but oral recording plays an important role in this process. Barber (2009: 4–5) notes how recordings of oral texts for broadcasting subtly transform the text into something else: "… conceptualized perhaps for the first time as a boundaried entity that can be described, documented and exemplified …" (2009: 4), and how through being recorded an oral text can assume different functions, such as the signalling of a tradition, claims to recognition, or affirmations of autonomy by a particular community. The effects of standardization, 'purification'

and speaker ideologies in the text genesis of our Kagulu example described above are very consistent with these prior observations.

Since the process of transcription and editing is influenced by the speakers' perception of written Kagulu and their literacy ideologies, it also reflects differences in education and language attitudes, and the different roles of the narrator, transcriber and editor of the story. The story was narrated by an older woman without literacy training, while the editor was in his thirties and educated to university level. Both were brought up in the heartland of the Kagulu area with two Kagulu-speaking parents. Although the narrator lived in a Kagulu village, she worked in a Swahili-speaking setting, while the editor moved to Morogoro town, a non-Kagulu speaking area. The editor was married to another Kagulu speaker and spoke Kagulu at home on a daily basis. He was a Bible translator at the time, and likely to be more aware of Swahili influence and of efforts to preserve his language, and as a result more prone to using ancestral code and seeing Kagulu structures as more authentic. As a translator and member of the Kagulu Bible translation review board, the editor was versed in textual editing and translation, as well as a central participant in Kagulu language activism. His use of literary and pure Kagulu must be seen against this background. It is noteworthy that the more mixed, original text (Text 1) was produced by an older speaker, in her sixties at the time of recording, compared to the transcriber and editor who were in their thirties. The use of a 'purer' form of Kagulu, showing less influence from Swahili, is thus associated in this example with younger speakers, contrary to what has often been reported in other situations. The example shows that the question of language use and age is here embedded into the specific sociolinguistic background of the members of the documentation project – such as education and training, rural vs. urban residence, language attitudes and literacy ideologies, and their specific function in the project.

There are also differences of audience and an associated change in the communicative and pragmatic context. The recorded text was narrated to a very small audience present at the event, although it was also recorded, while the edited text was aimed at potential future audiences of a corpus of the language. Being part of a description project, both texts were also aimed in a more immediate sense at the researcher. In the case of the narration, the researcher was present, which may have affected the narrator, while the editor took the texts home and worked on his own, thus being able to reflect more deeply on the intended audiences and the appropriate language forms for them.

The comparative study of the different versions of the text presented here shows, on the one hand, speakers' awareness of features of 'mixed' and 'pure' Kagulu, in particular forms borrowed from Swahili. On the other hand, the study shows that for this particular text form – a story intended to be part of a descriptive corpus of the language – speakers (or at least the transcriber and the editor) felt that 'pure' Kagulu was more appropriate.

The example is of interest for language documentation in particular because it shows the dynamics of language contact and variation, and how speakers manipulate the different forms and codes available in contact situations. From the point of view of linguistic description, none of the three variants of the text (the original narration, the transcribed version and the edited text) is more valid than any other. The recorded version is probably a more faithful record of how the language is used, and as such is a more 'authentic' representation of contemporary linguistic practice. On the other hand, the edited version includes forms probably associated with a more reflective register, and is an example of what the editor

conceives as being a purer and in some sense more 'authentic' form of the language. It also includes lexical items such as *nhambo* 'journey' and *kamei* 'then', which, if restricted to such a 'reflective' register, would be lost if only the recorded text was available.

Investigating text genesis in detail, as we have done here, thus also raises questions about the status of existing documents in language documentation. For example, Hult and Wahlgren (2012) describe similar effects in the construction of classical Greek texts and note that only some versions survive for posterity and documentation. For Kagulu, a number of published texts exist in the form of transcribed oral stories (Beidelmann 1967, 1971, 1997 a.o.). Beidelman (1997: xv) notes the variation inherent in the texts, and how transcription needs to be faithful to variation, as "it may reflect speakers' language competencies, indicate their ideas about their language and how it should be represented". However, compared to our example (and similar recordings in the Kagulu ELAR corpus), virtually no instances of variation of the kind seen in our Text 1 are to be found in the stories in Beidelman (1967) (see Petzell 2015). This might be due to the fact that the texts were transcribed directly – without an audio recording – by Kagulu speakers who "were literate in their own language in addition to Swahili and some English" (1997: xv). As our discussion has shown, this transcription process is likely to have added its own dynamic, placing the texts in a particular context of written textual development – and so their linguistic interpretation will be helped by a more comprehensive understanding of this dynamic.

From a sociolinguistic perspective, the process of editing itself, and the effects of the editing process, rather than the eventual product, are thus of central importance. What the process shows is the establishment of a version of the text which can be seen as a more traditional and conservative variety, aiming to preserve an ancestral code (Woodbury 2005, 2011). The addition of Kagulu opening and closing formulas (the closing formula *simo ihela* replacing the corresponding Swahili-influenced version *hadithi imeisha*) establishes the text more firmly as an authentic Kagulu story (cf. Mosel 2008, 2012). The replacement of the high-frequency adverb *halafu* with *kamei* has a strong effect on the text overall by significantly reducing the token frequency of Swahili-influenced forms. The overall effect of replacing 'mixed' Kagulu forms with 'pure' Kagulu forms is to make the text appear more distant from Swahili, thus representing a more distinct and unique variety. The result is a 'purer' version of Kagulu, but one which might in fact be a more homogeneous and more essentialized version of the language than is actually used, or has been used by speakers in the past.

In sum, the example shows the interdependence of the different versions of the text. One version without the others would present only part of the complex Kagulu situation, and would show only one linguistic reality at the expense of others. In contrast, by looking at different versions, and the contexts and processes of their production and perception, we can develop a better and richer understanding, not only of the Kagulu language, but also of the specific multilingual dynamics of language contact and linguistic variation in which its speakers are embedded.

4. Conclusions. In this paper we have presented an example of linguistic variation, language contact and the development of texts within the dynamics of language documentation. We have noted that many Tanzanian ethnic community languages show lexical and structural influence from Swahili, and that speakers often distinguish between 'pure' and

'mixed' varieties, differentiated by more or less influence from Swahili, and often associated with younger vs. older speakers.

In the case of Kagulu, we have shown that influences from Swahili are found at the lexical and grammatical level, and that in many cases, variant forms result from the adaptation of a Swahili form. However, while a list of variant forms provides a static picture of contact effects, it provides only limited information about the usage of different forms. Since language documentation aims to capture variation within a given language, as well as the way speakers interact with and use this variation, we have drawn attention to the dynamic aspects of language documentation by presenting a case study of three different versions of a Kagulu story – a recorded oral version, its transcription, and an edited subsequent version of it – and the dynamics between the three versions.

The comparison has shown the process of editing in 'pure' Kagulu, through the replacement of Swahili-influenced forms, the addition of forms which are felt to be more authentically Kagulu and the addition of forms that reinforce the genre of the text. This process is influenced by various factors, including a change of genre from spoken to written text, as well as by the complex sociolinguistic background of the speakers involved in the text genesis – the narrator, the transcriber and the editor – and the differences in their language use, educational background and professional training, age and residence, and language attitudes and literacy ideologies.

We have proposed that all versions of the text are valid in their own right, and that they are 'authentic' representations of different linguistic usage. Each text documents its own linguistic reality, and without any one version, our understanding of Kagulu would be poorer and more partial. Furthermore, a comparison of the texts, and the processes which led to their existence, provides a handle on the process of adjustment and editing which relates the versions to each other, thus producing a richer description of the language and of how speakers use and exploit patterns of variation for the construction of specific versions of the language. The example shows the interaction of linguistic variation, language contact and linguistic ideologies, and the relevance and importance of these dynamics for language description and documentation.

REFERENCES

Barber, Karin. 2009. Orality, the media and new popular cultures in Africa. In Kimani Njogu & John Middleton (eds.), *Media and identity in Africa*, 3-18. Edinburgh: Edinburgh University Press for the International African Institute.

Batibo, Herman M. 1992. The fate of ethnic languages in Tanzania. In Matthias Brenzinger (ed.), *Language death: Factual and theoretical explorations with special reference to East Africa*, 85-98. Berlin: Mouton de Gruyter.

Batibo, Herman M. 2005. *Language decline and death in Africa: Causes, consequences and challenges.* Clevedon: Multilingual Matters.

Beidelman, Thomas O. 1967. Eleven Kaguru tales. *African Studies* 26(1). 3-36.

Beidelman, Thomas O. 1971. *The Kaguru: A matrilineal people of East Africa.* New York: Holt, Rinehart & Winston.

Beidelman, Thomas O. 1997. *The cool knife: Imagery of gender, sexuality, and moral educa-*

tion in Kaguru initiation ritual. Washington D.C.: Smithsonian Institution Press.

Childs, Tucker, Jeff Good & Alice Mitchell. 2013. Beyond the ancestral code: Towards a model for sociolinguistic language documentation. Paper presented at ICLDC 3, Honolulu, 3 March 2013.

Croft, William. 2003. *Typology and universals.* 2nd edn. Cambridge: Cambridge University Press.

Duranti, Alessandro. 1997. *Linguistic anthropology.* Cambridge: Cambridge University Press.

Geider, Thomas. 2003. *Motivforschung in Volkserzählungen der Kanuri (Tschadsee-Region).* Cologne: Rüdiger Köppe Verlag.

Guthrie, Malcolm. 1967-71. *Comparative Bantu.* 4 vols. Farnborough: Gregg.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161-195.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics. Studies and Monographs, 178), 1-30. Berlin: Mouton de Gruyter.

Hinton, Leanne. 2001. New writing systems. In Leanne Hinton & Ken Hale (eds.), *The green book of language revitalization in practice*, 239-250. San Diego: Academic Press.

Hult, Karin & Staffan Wahlgren. 2012. Traderingen av grekiska texter och grekisk textkritik. [Transmission of Greek texts and Greek textual criticism.] In Karin Hult, Christina Thomsen Thörnqvist & Gunhild Vidén (eds.), *Hunden och lopporna: uppsatser om editionsfilologi*, 30-41. Stockholm: Svenska vitterhetssamfundet.

Kahigi, Kulikoyela, Yared Kihore & Maarten Mous (eds.). 2000. *Lugha za Tanzania – Languages of Tanzania.* Leiden, The Netherlands: Universiteit Leiden; Research School for Asian, African, and Amerindian Studies (CNWS).

Kiango, John G. 2013. Swahilization of the coastal Bantu languages: The case of the Bondei language. In Karsten Legère (ed.), *Bantu languages and linguistics: Papers in memory of Dr Rugatiri D. K. Mekacha*, 199-208. Bayreuth: Bayreuth African Studies Series.

Last, J. T. 1886. *Grammar of the Kaguru language, eastern equatorial Africa.* London: Society for Promoting Christian Knowledge (SPCK).

Legère, Karsten. 1992. Language shift in Tanzania. In Matthias Brenzinger (ed.), *Language death: Factual and theoretical explorations with special reference to East Africa*, 99-115. Berlin: Mouton de Gruyter.

Legère, Karsten. 2002. The 'Languages of Tanzania' project: Background, resources and perspectives. *Africa and Asia* 2. 163-186.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the world.* 17th edn. http://www.ethnologue.com.

Maho, Jouni F. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 639-651. London: Routledge.

Maho, Jouni F. & Bonny Sands. 2002. *The languages of Tanzania: A bibliography.* Gothenburg: Acta Universitatis Gothoburgensis.

Marten, Lutz. 2013. Linguistic variation, language contact and the new comparative Bantu. Inaugural lecture, SOAS, University of London, 18 April 2013. https://www.youtube.

com/watch?v=QDrZRZTCcyU

Mbaabu, Ireri. 1985. *New horizons in Kiswahili: A synthesis in developments, research and literature.* Nairobi: Kenya Literature Bureau.

Meeusen, Achille Emile. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 80-121.

Mekacha, R.D.K. 1993. *The sociolinguistic impact of Kiswahili on ethnic community languages in Tanzania: A case study of Ekinata.* Bayreuth: Bayreuth University Press.

MLT (Mradi wa Lugha za Tanzania; Languages of Tanzania Project). 2009. *Atlasi ya lugha za Tanzania.* Dar es Salaam: Mradi wa Lugha za Tanzania, Chuo Kikuu cha Dar es Salaam.

Mosel, Ulrike. 2008. Putting oral narratives into writing – experiences from a language documentation project in Bougainville, Papua New Guinea. Paper presented at Simposio Internacional Contacto de lenguas y documentación. Buenos Aires, 14-15 August 2008. http://www.linguistik.uni-kiel.de/Oral%20narratives%20into%20writing.pdf

Mosel, Ulrike. 2012. Creating educational materials in language documentation projects – creating innovative resources for linguistic research. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek (eds.), *Potentials of language documentation: Methods, analyses, and utilization*, 118-125. Honolulu: University of Hawai'i Press.

Muzale, Henry R.T. & Josephat M. Rugemalira. 2008. Researching and documenting the languages of Tanzania. *Language Documentation and Preservation* 2(1). 68-108.

Nicolle, Steve. 2013. *A grammar of Digo: A Bantu language of Kenya and Tanzania.* Dallas, Texas: SIL International.

Ochs, Elinor. 1979. Transcription as theory. In Elinor Ochs & Bambi B. Schieffelin (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.

Petzell, Malin. 2008. *The Kagulu language of Tanzania: Grammar, texts and vocabulary.* Cologne: Rüdiger Köppe Verlag.

Petzell, Malin. 2012a. The linguistic situation in Tanzania. *Moderna Språk* 106(1). 136-144.

Petzell, Malin. 2012b. The under-described languages of Morogoro: A sociolinguistic survey. *South African Journal of African Languages* 32(1). 17-26.

Petzell, Malin. 2015. Three Kagulu stories: Annotations, analysis, and word lists. *Studia Orientalia Electronica* 3. 48-95.

Rugemalira, Josephat M. 2010. The *-ag-* TAM marker and the boundary between cliticization and inflection in Bantu. In Karsten Legère & Christina Thornell (eds.), *Bantu languages: Analyses, description and theory*, 229–237. Cologne: Rüdiger Köppe Verlag.

Slabbert, Sarah, & Rosalie Finlayson. 1998. Comparing Nguni and Sotho: A sociolinguistic classification. In Ian Maddieson & Thomas J. Hinnebusch (eds.), *Language history and linguistic description in Africa*, 288-306. Trenton, NJ: Africa World Press.

Whiteley, W.H. 1969. *Swahili: The rise of a national language.* London: Methuen.

Woodbury, Anthony C. 2005. Ancestral languages and (imagined) creolization. *Language Documentation and Description* 3. 252-262.

Woodbury, Anthony C. 2011. Language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 159-186. Cambridge: Cambridge University Press.

Yoneda, Nobuko. 1996. The impact of the diffusion of Kiswahili on ethnic languages in Tan-

zania: A case study of Samatengo. In S. Hino (ed.), *African urban studies IV*, 29-73. Tokyo: Research Institute for the Study of Languages and Cultures of Asia and Africa.

Yoneda, Nobuku. 2010. "Swahilization" of ethnic languages in Tanzania: the case of Matengo. *African Study Monographs* 31(3). 139-148.

APPENDIX

'The hyena and the rabbit', recorded on 28 Aug 2003 by Malin Petzell; ELAR Kagulu deposit (http://elar.soas.ac.uk/deposit/petzell2010kagulu); original transcription mjs1.doc, sound file mjs1.waf; re-transcribed in May 2013 by Lutz Marten, Nancy C. Kula and Malin Petzell.

The text is presented in five lines:
1) Original recording (retranscribed) (Text 1/5)
2) Edited text (Text 4)
3) Morpheme-by-morpheme analysis
4) Morphological glosses
5) Parts of speech

Differences between lines 1 and 2 (i.e. between Text (1/5) and Text (4)) are highlighted by bold typeface and grey shading. A free, idiomatic translation is included after each sentence, and a translation of the whole story is provided before the text. Abbreviations follow the Leipzig glossing rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php) and Croft (2003), with some additions that have no equivalence in the glossing rules.[8] The morpheme-by-morpheme analysis represents underlying forms which do not always correspond to the surface form in the text lines. For instance, the NCP of class 5, *di-*, is often realized as *i-*. The glosses in the gloss line are literal, while the free translation below each segment is idiomatic; the latter may, therefore, differ slightly from the glossing. Noun class membership (gender) is indicated after each noun, so that 'hare:9/10' denotes that the singular of hare is in class 9 and the plural in class 10.

THE HYENA AND THE RABBIT

A long time ago, the hyena and the rabbit were friends. Then one day, the rabbit told the hyena, "Let's go on a journey." They arrived at a house and were welcomed. They stayed there for many days. One day the rabbit said to the owner of the house, "Tomorrow we are going back home, okay?" At night the rabbit ate all the peanuts, and afterwards he took some water into his mouth, swilled it around his mouth, and went to spit over the hyena. In the morning of the following day, the rabbit told their host, the owner of the house, "Check

---

[8] Abbreviations that are not in the Leipzig glossing rules: 1, 2, 3 = noun class number, ACP = agreement class prefix, CONJ = conjunction, EXT = extension, FV = final vowel, INT = interrogative, INTERJ = interjection, PP = pre-prefix, N = noun, NEUT = neuter, NCP = nominal class prefix, NUM = numeral, OM = object marker, PRN = pronoun, REF = referential, SM = subject marker, TM = tense marker, V = verb.

your belongings to see if they are okay." So the owner of the house checked his belongings, and he found that his peanuts had been eaten. Then the rabbit told him, "Bring a chair, and let's wash our mouths. Each of us shall gargle so we can see who stole and ate the nuts." The rabbit himself went first, saying, "Give me the water so that I can go first and gargle." Then he washed out his mouth, and spat clean water onto the chair. Next it was the hyena's turn. He washed out his mouth and was found to be the one who had eaten. The hyena was killed. That is the end of the story.

THE HYENA AND THE RABBIT

		difisi	na	sungula
Baho	**katali**	difisi	na	sungula
baho	katali	di-fisi	na	sungula
DEM	long time ago	5-hyena:5/6	CONJ	hare:9/10
prn	*adv*	*ncp-n*	*conj*	*n*

hawowa	mbuya	**ha(l)afu**	
hawowa	mbuya	**kamei**	
ha-wa-uw-a	mbuya	kamei	
PST-2-be-FV	friend:1a,9/2,10	then	
tm-sm-v-fv	*n*	*adv*	

sungula	kamgamba,	"chigende	**safali."**
sungula	kamgamba,	"chigende	**nhambo."**
sungula	ka-m-gamb -a	chi-gend-e	N-tambo
hare:9/10	1.PST-1-speak-FV	7-go-FV	9/10-journey:9/10
n	*sm-ncp-v-fv*	*sm-v-fv*	*ncp-n*

A long time ago, the hyena and rabbit were friends, then rabbit told the hyena "Let us go on a journey."

Ha(l)afu	howoluta	kunyumbangwa	**imonga,**
Kamei	howoluta	kunyumbangwa	**imwe,**
kamei	ha-wa-lut -a	ku-nyumba-ngwa	i-mwe
then	PST-2-go-FV	17-house:9/10-somebody's	9-one
adv	*tm-sm-v-fv*	*ncp-n-prn*	*ncp-num*

wakalibishigwa.
wahokeligwa.
wa-hokel-igw-a
2-receive-PASS-FV
sm-v-tm-fv

Then they went to a house and they were welcomed.

Ha(l)afu	**wekala**	bahaya	majuwa	mengi
Kamei	**howekala**	bahaya	majuwa	mengi
kamei	ha-wa-ikal-a	ba-ha-ya	ma-jua	ma-ingi
then	PST-2-stay-FV	LOC-16-DEM	6-day:5/6	6-many
adv	*tm-sm-v-fv*	*acp-ncp-prn*	*ncp-n*	*ncp-quant*

They stayed there for many days.

Ha(l)afu

Ijuwa	**dimonga**	sungula	**kamgamba**
Ijuwa	**dimwedu**	sungula	**hoyomgamba**
di-jua	di-mwe(du)	sungula	ha-ya-o-m-gamb-a
5-day:5/6	5-one	hare:9/10	PST-1-REF-1-speak-FV
ncp-n	*ncp-num*	*n*	*tm-sm-ref-om-v-fv*

		"aseye	nhosiku
mwinya	**ikaya**	"aseye	nhosiku
mu-inya	i-kaya	a-seye	nhosiku
1-having	PP-village:9/10	PP-1PL	tomorrow
ncp-prn	*iv-n*	*iv-prn*	*n*

chikuluta	ukaya,	haya!"
chikuluta	ukaya,	haya!"
chi-ku-lut-a	u-kaya	heye
7-FUT-go-FV	14-home:14	ok
sm-tm-v-fv	*ncp-n*	*interj*

One day the rabbit said to the owner of the house "Tomorrow we are going back to our home, ok!"

yakumtamila mwenye nyumba, halafu

Nhechilo	sungula	(kadiya	mayowe)x2	gose,
Nhechilo	sungula	kadiya	mayowe	gose,
nhechilo	sungula	ka-diy-a	ma-yowe	ga-ose
at.night	hare:9/10	1.PST-eat-FV	6-peanut:5/6	6-all
adv	*n*	*sm-v-fv*	*ncp-n*	*acp-prn*

harafu	kalonda	meji,	kasugusa,
kamei	kalonda	meji,	kasugusa,
kamei	ka-lond-a	meji	ka-sugus-a
then	1.PST-take-FV	water:6	1.PST-gargle-FV
adv	*sm-v-fv*	*n*	*sm-v-fv*

keja	mutwila	difisi.
keja	mutwila	difisi.
ka-ij-a	m-tw-il-a	di-fisi
1.PST-come-FV	1-spit-APPL-FV	5-hyena:5/6
sm-v-fv	*om-v-ext-fv*	*ncp-n*

At night the rabbit ate all the peanuts, then he took water into his mouth, washed his mouth, and went to spit over the hyena.

Harafu	nhosikusiku	usiku uja	ha(r)afu
Kamei	nhosikusiku		
kamei	nhosikusiku		
then	morning:9		
adv	*n*		

yakumtamila
hoyomtamila
ha-ya-o-m-tamil-a
PST-1-REF-1-explain-FV
tm-sm-ref-om-v-fv

	mwenye	nyumba,
imukaya	**mwinya**	nyumba,
i-mu-kaya	mu-inya	nyumba
PP-18-house:9/10	3-having	house:9/10
iv-ncp-n	*ncp-prn*	*n*

"ulangise	finhu	fyako
"ulangise	finhu	fyako
u-langis-e	fi-nhu	fi-ako
2SG-search-FV	8-thing:7/8	8-2SG.POSS
sm-v-fv	*ncp-n*	*ncp-prn*

halika	fyose	fiswamu."
halika	fyose	fiswamu."
ng'halika	fi-ose	fi-swamu
if	8-all	8-good
adv	*acp-prn*	*ncp-adj*

In the morning of the following day, the rabbit told their host, the owner of the house, "Check your belongings to see if they are all right."

Ha(r)afu	mwenye	nyumba	kalangisa
Kamei	imwinya	kaya	kalangisa
kamei	i-mu-inya	kaya	ka-langis-a
then	PP-1-having	house:9/10	1.PST-search-FV
adv	*iv-ncp-prn*	*n*	*sm-v-fv*

finhu	fyakwe,	kona	mayowe
finhu	fyakwe,	kona	mayowe
fi-nhu	fi-akwe	ka-on-a	ma-yowe
8-thing:7/8	8-3SG.POSS	1.PST-see-FV	6-peanut:5/6
ncp-n	*acp-prn*	*sm-v-fv*	*ncp-n*

gakwe	gadiigwa,	harafu	
gakwe	gadiigwa,	kamei	sungula.
ga-akwe	ga-diy-igw-a	kamei	sungula
6-3SG.POSS	6-eat-PASS-FV	then	hare:9/10
acp-prn	*sm-v-tm-fv*	*adv*	*n*

aka	kamgamba,	"lete	digoda
	kamgamba,	"lete	digoda
	ka-m-gamb-a	let-e	di-goda
	1.PST-1-speak-FV	bring-FV	5-chair:5/6
	sm-om-v-fv	*v-fv*	*ncp-n*

chisuguse,	kila	munhu	yasuguse
chisuguse,	kila	munhu	yasuguse
chi-sugus-e	kila	mu-nhu	ya-sugus-e
1PL-gargle-FV	each	1-person:1/2	1-gargle-FV
sm-v-fv	*adj*	*ncp-n*	*sm-v-fv*

yoneke	yahijile		yadiile
yoneke	yahijile	au	yadiile
ya-on-ek-e	ya-hij-ile	au	ya-diy-ile
1-see-NEUT-FV	1-steal-PFV	or	1-eat-PFV
sm-v-ext-fv	*sm-v-tm*	*conj*	*sm-v-tm*

nani	no	yehoki".
	no	ya-hoki
who	COP	1-INT
prn	*cop*	*acp-int*

Then the owner of the house checked his belongings, and he found that his peanuts were eaten, then rabbit told him, "Bring a chair so that we can wash out our mouths, each person will gargle so one can see who stole or who ate."

"ning'heni	nikonge	niye	kusugusa	meji."
"ning'heni	nikonge	niye	kusugusa	meji."
ni-ing'h-e-ni	ni-kong-e	niye	ku-sugus-a	meji
1SG-give-FV-PL	1SG-start-FV	1SG	15-gargle-FV	water:6
sm-v-fv-ext	*sm-v-fv*	*prn*	*ncp-v-fv*	*n*

Then rabbit himself started, "Give me the water so that I can gargle first."

Ha(r)afu	kasugusa	meji,	kotwila
Kamei	kasugusa	meji,	kotwila
Kamei	ka-sugus-a	meji	ka-tw-il-a
then	1.PST-gargle-FV	water:6	1.PST-spit-APPL-FV
adv	*sm-v-fv*	*n*	*sm-v-ext-fv*

mwigoda	meji	maswanu;	**ha(r)afu**	keja
mwigoda	meji	maswanu;	**kamei**	keja
mu-di-goda	meji	ma-swanu	kamei	ka-ij-a
18-5-chair:5/6	water:6	6-good	then	1.PST-come-FV
ncp-ncp-n	*n*	*ncp-adj*	*adv*	*sm-v-fv*

difisi,	kasugusa	koneka
difisi,	kasugusa	koneka
di-fisi	ka-sugus-a	ku-onek-a
5-hyena:5/6	1.PST-gargle-FV	15-find out-FV
ncp-n	*sm-v-fv*	*ncp-v-fv*

kifwa	imwene	noyadiile.	
fina	imwene	noyadiile.	
fina	i-mu-ene	no	ya-diy-ile
that	PP-1-PRN	COP	1-eat-PFV
conj	*iv-ncp-prn*	*cop*	*sm-v-tm*

Then he washed out his mouth and spat clean water onto the chair; then the hyena came, washed out his mouth and was found to be the one who had eaten.

Hadithi imeisha.

Difisi	dikomigwa.		
Difisi	dikomigwa.	**Simo**	**ihela.**
di-fisi	di-kom-igw-a	simo	i-hel-a
5-hyena:5/6	5-kill-PASS-FV	story:9	9-finish-FV
ncp-n	*sm-v-tm-fv*	*n*	*sm-v-fv*

The hyena was killed. The story is over.

Lutz Marten
lm5@soas.ac.uk

Malin Petzell
malin.petzell@sprak.gu.se