TILBURG UNIVERSITY
Understanding Society

Tilburg Papers in Culture Studies

Paper 34

Comparing spelling and segmentation practices in three versions of a Mandinka text

by

Kasper Juffermans & Dorina Veldhuis

kasper.juffermans@uni.lu
t.m.veldhuis@uvt.nl

September 2012

Comparing spelling and segmentation practices in three versions of a Mandinka text

by

Kasper Juffermans and Dorina Veldhuis

kasper.juffermans@uni.lu

t.m.veldhuis@uvt.nl

September 2012

Abstract

In this paper combines insights, data and methods from two projects conducted in two different research traditions: an ethnographic sociolinguistic study of literacy products in a West African society (The Gambia), and an experimental cognitive linguistic study into the influence of literacy on spelling practices and segmentation of linguistic units. Our paper reviews experimental research into non-literates' metalinguistic awareness and analyses texts from the ethnographic study in order to address questions of units of language and units of writing in Mandinka. Through a comparative text analysis of three differently authored versions of a short text in Mandinka, ‘the donkey story’, we argue that awareness of units in language not only depends on the language, script or writing system in which one acquires literacy, but that the sociolinguistic context and the orthographic regime of the society in which people learn to write, spell and segment, also matters.

Keywords

Literacy practices; grassroots literacy; units of language and writing; metalinguistic awareness; spelling

Spelling and segmentation practices across three versions of a Mandinka text: A comparative analysis

1 Introduction

Throughout history, linguists have addressed the question what the basic units for the perception and processing of human languages are. Sapir (1970 [1921]), for instance, suggested that words are the basic units of language. Later work in the psychology of language argued that sub-word units like syllables or morphemes are the units children first become aware of when they acquire language (cf. Berthoud-Papandropoulou, 1978) and more recently, cognitive linguistics has argued that multi-word constructions play a major role in child language acquisition and language processing (cf. Tomasello, 2006; Arnon, 2010). For those educated in the Western world and making daily use of alphabetic writing systems in which words are the basic units, separated by spaces, it is a powerful idea that language is naturally segmented in words. Also in studies with a generative approach, words and sentences are often still taken as natural units of language and as the basis for language research (e.g., Pinker, 2007; Chomsky, 1965). The question if words or something smaller or larger than words are the basic units for the perception of language has been topic of discussion in studies with pre-literate children and non-literate adults (e.g., Karmiloff-Smith, Grant, Sims, Jones and Cuckle, 1996; Kurvers and Uri, 2006) and cross-linguistic studies on metalinguistic awareness of speakers of languages with writing systems that do not mark words with spaces (Bassetti, 2005; Hoosain, 1992). These studies suggest that one's idea of words as units in language is influenced by the writing system.

In relation to this question, there have been studies into general effects of literacy on societies and members of mainly non-literate societies (Goody and Watt, 1963; Scribner and Cole, 1981) as well as studies concerned with orthography development for language (and communities) without acknowledged written traditions (Lüpke, 2011; Van Dyken and Kutsch Lojenga, 1993). In this paper we discuss writing and literacy in an African language and society in which literacy is not formally educated and not commonly practiced. Our paper is not concerned with general literacy effects on societies or with developing orthographies, but with

describing spelling and segmentation practices in different moments of entextualisation in order to shed light on the instability and artificiality of units of language and units of writing.

Our paper brings together insights, methods and data from two different linguistic subdisciplines in an attempt to provide a richer and fuller account of language and literacy learning outcomes with respect to spelling and segmentation practices. The paper is situated at the intersection of two larger projects carried out at Tilburg University, one an ethnographic sociolinguistic study of multilingual literacy products and practices in The Gambia, West Africa (see Juffermans 2010), the other an experimental cognitive linguistic study into metalinguistic awareness, language segmentation and the awareness of units (the 'building blocks') of language (see Veldhuis & Backus 2012; Veldhuis & Kurvers 2012). It combines insights from the experimental study – in which child segmented language data are taken as a basis for conclusions about language and cognition – with data and methods from the ethnographic sociolinguistic study – in which everyday literacy products and practices are described and analysed to arrive at conclusions about language and society. This combination of insights, methods and data is meant to provide a more complete picture of both cognitive and social aspects of Mandinka spelling and segmentation practices.

The remainder of this paper is divided in four main sections. We first briefly review psycholinguistic experimental research into word segmentation and metalinguistic awareness, followed by a description of the sociolinguistic context of Mandinka literacy in The Gambia. Hereafter, we embark on a comparison of spelling and segmentation practices in three versions of a short text in Mandinka, 'the donkey story': the original written by a non-formally educated peri-urban middle-aged man, and two respellings written by a formally educated younger urbanite. In our comparative analysis of the texts, we first discuss spelling differences across the three versions, followed by a discussion of the segmentations and word boundaries in the three versions.

2 Language segmentation and writing systems

Numerous studies on language segmentation are related to the development of children's metalinguistic awareness. Since the 1970s, many theoretical and experimental studies have been published on the way in which children segment spoken language, and on their development of the awareness of specific units in language, such as words and phonemes.

In older studies of segmentation and the processing units of language, it was argued that words are the basic or natural units of language (cf. Sapir, 1970 [1921]). According to this idea, no matter whether one is literate or not, and irrespective of one's native language, it was suggested that people universally use words as basic or natural units when processing and producing language (Kraak, 2006:132). Experimental studies into children's metalinguistic awareness and cross-linguistic studies have suggested that people's awareness of words and phonemes only develops with literacy in languages in which such units are marked in writing (Homer, 2000; Kurvers, 2002; Kurvers and Uri, 2006; Morais, Bertelsen, Cary and Alegria, 1986; Ramachandra and Karanth, 2007). Pre-literate children and non-literate adults have shown difficulties in segmenting phonemes or words from speech, whereas literate subjects who speak languages in which words and letters are marked in writing are generally able to do this. However, in studies by Hoosain (1992), Bassetti (2005) and Veldhuis et al. (2010) for instance, it was found that Chinese, who not mark words in their text lay-outs, have no univocal idea of what words are, and that Chinese native-speakers only become aware of the units that are marked in their writing system, characters, when they learn to read and write. This supports Bugarski's (1993) idea of 'graphic relativity', which claims that people using different languages in writing recognise different orthographic units.

In their classic study on the psychology of literacy, Scribner and Cole (1981) also discussed the influence of specific writing systems on people's perception. They showed among other things that their literate research subjects among the Vai people in Liberia, who make use of a syllabic writing system, approached research tasks such as solving rebuses differently from non-literate persons or persons who make use of alphabetic writing systems. Literate Vai were able to transfer their knowledge about decoding from their writing system to decoding a rebus, as they appeared to be better in such exercises than non-literate Vai and Vai who were literate

in English or Arabic. Accordingly, Scribner and Cole concluded that being literate in a specific writing system affects the way in which one approaches decoding exercises.

3 English, Mandinka and literacy practices in The Gambia

Let us now turn to the ethnographic sociolinguistic study of literacy and multilingualism in The Gambia to shed light on units of language and writing from a non-experimental perspective.

Our take on ethnography here is two-folded. First, the data used in this paper are ethnographic data as they have been collected and analysed as apart of a project investigating literacy and multilingualism in The Gambia, involving ethnographic fieldwork in the form of participant observation and "deep hanging out" over extended periods and a holistic and dynamic ethnographic research focus rather than a set of detailed research questions (see Juffermans, 2010 for details). This paper, however, is not an ethnographic study in that perspective, as the question we have set out to answer here is not answered following such an ethnographic methodology. This paper is ethnographic rather in its analysis of the texts, i.e. in its approaching of texts as indexically ordered products of human language practices, as connected to its production and uptake and as containing traces of social structure and power relations, of language and literacy ideologies including writers' linguistic intuitions and ideas about how language should be segmented into words. This paper is ethnographic in the sense of an "ethnography of text": in being a descriptive approach to writing in the Hymesian tradition of ethnography of speaking (Basso, 1974; Hymes, 1974) which was directed towards understanding literacy events and practices and their role and function in social life (Blommaert, 2008; see also Baynham, 2004 for a pedigree of ethnographic literacy studies).

Before we move to the analysis of texts for what they reveal about language and literacy practices in Gambian society, we will first have to sketch the social and linguistic context in which these texts are situated. This sketch is based on Juffermans and McGlynn (2009).

When measured in absolute terms of literacy vs. non-literacy, The Gambia exhibits rather low literacy rates. The overall literacy rate is, unspecified for age, estimated at 46% for both sexes and only 28% percent for women (DoSE, 2006: 44). The "problem", however, is more complex than these numbers suggest: literacy in a multilingual society such as The Gambia is not just about who can and who cannot write, but is equally about how and in what language those who "can write" write, and in what language they do not or cannot write.

Official publications usually cite the following nine ethnolinguistic groups in order of decreasing number of people: Mandinka, Fula, Wolof, Jola, Serahule, Serer, Manjago, Bambara and Aku. The Gambia's two most widely spoken languages, Mandinka and Wolof are typologically quite different and fall under two distinct language families, Mande and Atlantic languages respectively. Mandinka is most widely spoken as a first and second language in the rural areas up-country, and Wolof assumes the role of vehicular language in urban Kombo in the west of the country, including the capital Banjul and the largest conurbation Serrekunda. Most Gambians speak the language of their ethnic group, the regional lingua franca (Mandinka or Wolof) and, although more exclusively, English. Multilingualism is the rule in The Gambia, whether on national, regional, family, individual or utterance level. Monolingual villages and persons are rather exceptional in the overall picture. In Gambian society, virtually everybody is multilingual but hardly anybody is literate in all the languages of his or her repertoire.

As a former British colony (until 1965), the official language of The Gambia is English. It is the language of the offices but is also the official medium of instruction in schools at all levels, from nursery schools to university level. As a predominantly Muslim country, Arabic occupies an important position in the Gambian ecology as well, in particular for initial greetings and praying, as well as for religious education. Surrounded by Senegal and in proximity to other francophone countries, French also has some formal and informal function in The Gambia, albeit limited.

Whereas school attendance was very low very throughout the colonial period (Hughes and Perfect, 2006:27f), most children are now subject to formal schooling of a particular type. School-going children are taught to read and write in either the official, post-colonial language English (in the secular public or private school system) or in the Arabic language of Islam and the Arabic world (in the Islamic *madrassa* school system). Adults who have missed the opportunity too learn either English or Arabic in their youth are offered an alternative, non-formal education known as 'adult literacy classes' in which they are taught to read and write in their community's local language. So, children and adults in education are subjected to two radically different regimes of learning. Adult in adult literacy seems to be synonymous with "local" or "vernacular" as opposed to the "global", "international" or "official" literacy of English and Arabic.

Mandinka and other Gambian languages have official Latin alphabet spelling systems that are employed and prescribed by the Gambian government and international (mainly Christian) NGOs working in The Gambia on (adult) literacy education and development (Faye and Sillah, 1956; Sidibe, 1979; WEC International, 2010 [undated]). In addition, there is a Mandenkan

vehicular cross-border language commission that is concerned, among other things, with creating a unified written standard for the varieties of Manding languages that are separately known in the respective national contexts as Bambara, Malinke, Dyula, Mandinka, etc. Although harmonisation of West African languages has been on the agenda for several decades (see e.g., Diagne, 1978), its activities are very much still "in progress" and have so far had very little, if any, impact on Gambian society.

Although part of the broader debate on language and literacy in Africa, this can hardly count as context for Mandinka spelling and segmentation practices on the ground. At present, there are language materials and an orthography in use for Gambian Mandinka, but not for the larger Manding cluster. More importantly, neither the regional standard in the making, nor the existing national standard orthography, is popularly 'enregistered' (i.e., socially recognised and practically accepted as norm, Agha, 2005; see also Dong, 2010). They are simply unknown to and not practiced by most people. Mandinka has an 'official' or standard spelling only in the sense that it is used by government departments and NGOs, but the fact that it is not commonly practiced makes it official or standard only in theory, not in practice. The most powerful "sponsors of literacy" (Brandt, 1997) in The Gambia (i.e., the Department of State for Education, Islamic clergy, various educational stakeholders) are only marginally concerned with literacy development in local languages and invest few resources in the systematic teaching of reading and writing in local languages. Literacy education in local languages is confined to the loosely organised adult literacy classes which are organised and attended on a voluntary basis with varying degrees of intensity and varying over time and space. As a consequence, writers of a language such as Mandinka are left relatively free to decide how to spell and segment their language. Spelling Mandinka is for a large part a creative and heterographic affair, rather than a matter of *ortho*-graphy.

4 Three versions of the donkey story

4.1 The textual material

The textual material analysed for this contribution was part of a larger study on literacy products and practices in Gambian society (Juffermans, 2010). They consisted of samples of "grassroots literacy" collected in the "linguistic landscape" (Gorter, 2006), schools and everyday life. "Grassroots literacy" is sub-elite literacy, common in postcolonial and developing world

contexts, that is produced at a distance from, or in the absence of, institutions of prescriptive elite-linguistic normativity (secondary schools, libraries, language academies and other institutions providing models for good language use) and under conditions of poor material and infrastructural support for writing (e.g., without dictionaries, spell-checkers, reference material, Google and often without a standard orthography at hand) (Blommaert, 2008; Fabian, 2001). Most of the texts encountered in both urban and rural spaces, in public and private domains in The Gambia were either in English or Arabic. Written language or literacy in local languages, as the texts in Mandinka discussed here, proved to be extremely hard to find.

The first text discussed here was written by a person called Burama Janne, a middle-aged man living with his two wives and nine children in the town of Farato near Brikama in the southwest of the country. Burama had not had any formal education and was largely self-educated aided by a short series of adult literacy classes he participated in as a young adult. He explained that he owed his former job as a travelling hide merchant in great part to his ability to keep records of his transactions on paper, but also wrote short stories and travel notes for his personal satisfaction. Burama was clearly proud of his ability to write Mandinka and pleased with the researcher's interest in his literacy practices and presented him with a number of his writings on paper but also on the wall of his house. One of the texts he thus made available was the text reproduced below in Fragment 1, a story about a donkey that we will refer to as 'the donkey story'. It is a story of an incident he experienced on the road during one of his travels to the town of Basse in the far east of The Gambia.

Fragment 1. Transcription and translation of Burama's version of the donkey story

1	dANKI 1985 danKI KESSI	*Donkey 1985 Donkey Problem*
	NNa FALoo FILITA AGa AYINI	*My donkey got lost. I searched and*
	nyINI FO MNATA AJe MODoo bulu	*searched for it. I found it on Modou's hand.*
	AYAA ASSITI SAREe TOOBALA	*He tied it on a [donkey] cart.*
5	KABIRIN ngana FALoo JE dooRon	*When I saw my donkey, always*
	ngaa ASUute KAATO nga suuTE-	*I recognised it, because I put a mark*
	RENGO le KE. FALoo . BALA . KARINne	*on the donkey. Then [I said]*
	SALAMA le KUN AKO MALe KUN SALAMU	*salaam-maleikum, he said maleikum-salaam.*
	NKO NNa FOloo le MU nyINTI, ATEKO	*I said this here is my donkey. He said*

10	HANII. nyIN MANKE . E lA FAlOOTI. nga	*no, this is not your donkey. We*
	nyuu , SAbAn ,SAnbAng . FO NTATA	*argued and argued until we went to*
	POlICE . Nganna KUMO . SAATA . ETe.fana	*the police. I told my story. He too*
	YAAlA KUMO-SAatA . POlICOlo. KO	*told his story. The police asked:*
	MUN TAMAn sere JUma le be . ElA FAlo ba la	*what mark, where is it on your donkey?*
15	YAA FO NTE FAnaa ngaa FO	*He said [his] and I also said [mine].*
	POlISO KO MO WO MO EYe TAMAn SERO	*The police asked both of us, your mark,*
	MENTO ELA FAloo BALA nIN-WO-	*where is it on your donkey? If it*
	MAN TARAJee NbE SORONNA	*does not match, I will lock you up.*
	POlISOlu KO TAA FA loo -	*The police said, go bring the donkey*
20	KAMAneng EYA AKOroce NTE MEN-	*and let me observe that where*
	TA MU FAlOOTI ngaMen FO JEE	*on the donkey I was described to see [the mark],*
	WOle naaTa TARA JE. keedIngo	*is indeed what we see on it. The man*
	ko Abe dIyAAmulA doRon polICOlu	*wanted to talk immediately. The police slapped*
	YA ATule bun EYA FAYI SEli, WOTO-	*him and put him in the cell. Therefore,*
25	WO mo WO ENyAN ElA FEngo	*everybody must recognize his property.*
	SUUTE lAAle	
27	ngIne mu KIITIYOOTI	*This is a court.*

An electronic transcription of the photographed document was prepared in the field, but the translation and annotation of the documents was further worked on at home in interaction with Mandinka-speaking informants over Skype and e-mail after and in between fieldwork periods. In the course of working with Burama's texts, the first transcription of the donkey story was given to another person, Dembo – an educated urbanite in his late thirties, who lived in the Greater Serrekunda, The Gambia's largest urban concentration. Dembo was educated up to senior secondary school level in the formal English-medium school system and was fluent in English but had not in his educational trajectory learned to read and write his native Mandinka, and did not usually write in any language other than English. He returned the photographed text a few days later, however not with an English translation as requested, but with a Mandinka respelling of the original. This respelling is reproduced in transcription in Fragment 2 below.

Fragment 2. Transcription of Dembo's handwritten respelling of the donkey story

28 NAA – FALOO FEE LEE TAH

~~FOR~~ – ~~NAA~~ NYAA NYENE FOR

30 N^{N}ATA A^{R}JEH MODOU BULUU

~~NY~~ A^{R}YAA SEETE^{H} SARE-TOO BALAA

KABERING NGANA FALOO JEH

DORONG NGA A^{R}SUTAY, KHA˙TUNG

NGA SUTAY-RANGO LEKEH FALOO-BALA.

35 KARIM.KOO– SALAMU-ALAY-KUM.

NKO – MALAY-KUM-SALAM.

NKO – NA^{R} FALOO – LEMU NYING-TEE^{H}

A^{R}-THE-KOO HANEE NYING MANG-KEH

ELA FALOO TEE^{H} – NGA NYONG SABANG

40 SAABANG FOR ~~AR~~ ~~NAT~~ NTATA POLICO.

NYANG-NA- KUMOO SAATA. A^{R}TEH FANANG

YAALA KUMOO SANTA. POLICO KO MUNG-TAA

MANG-SERR JUMAA LEH–BEH ELA FALOO BALA

44 AYAA-FOO-. NTEH FANANG NGAA FOO.

The respellling was accompanied with the comment that this was the "correct writing" and that "the first writing [i.e., Burama's version] had some mistakes". Dembo even remarked that Burama's text reflected a writing style of someone who "[does]n't know the language very well". Translation of the text, or rather a paraphrase, was only given orally. Dembo's claim that his spelling was a more correct way of writing Mandinka than Burama's is an appreciation of his own over Burama's writing that is based more on sociolinguistic than on linguistic "analysis", i.e. on his recognition of Burama – whom he did not know personally – as a non-formally educated writer. Dembo thus presents himself to the researcher as higher educated and more sophisticated and a better informant for learning (about) Mandinka than Burama.

Following a second request from the researcher for translation over e-mail several months later after returning home from the field, Dembo presented a second respelling together with a translation as reproduced in Fragment 3 below.

Fragment 3. Dembo's second, e-mailed respelling of the donkey story

NAA FALOO FEELE TAH,,,,,,i lost my donkey
NGA NYENEE FOR NNTATA A JEH MODOU BULU,,,,,i search for and i found it with modou
AYA SEETEH SARETOO BALA KABERING NGA FALOO JEH,,,,,,he tied the donkey on a cart when when i saw it
DORONG NGA SUTAY,,,,i recognise it instantly
KHATUNG NGA SUTAY RANGO LEEKEH FALOO BALA,,,,because i put an identification mark on the donkey
KAREM KO ASALAMU ALAY KUM,,,Karem said peace be upon you thats greetings
NKO MALAY KUM SALAM,,,,i said peace be upon you too
NKO NAFALO LEMU NYING TEE,,,,,i said these donkey is mine
ARTEH KO HANEE NYING MANG KEH ELA FALOO TEE,,,,he said theses not your donkey
NGA NYONG SABANG SABANG JEH,,,,we argued
FOR NTATA POLICE,,,,untill wereach police
NGANA KUMOO SAATA ,,,,,,i gave my statement
ARTEH FANANG YAALA KUMOO SATA,,,he also gave his statement
POLICO KO MUNG TAMANGSERR JUMA LEBEH ELA FALOO BALA,,,the police asked as to wether you put an identification mark on your donkey
AYAA FOR NNTEH FANANG NGA FOR,,,,He explained and i explained too
END
In brief it was a missing donkey that was found in modous position of which he claimed ownership and it spark arguement of which they cannot settle and they finally have to go to the police and the police ask as to wether any of them did put an identity on the donkey and they all explained

4.2 A comparison of spelling practices

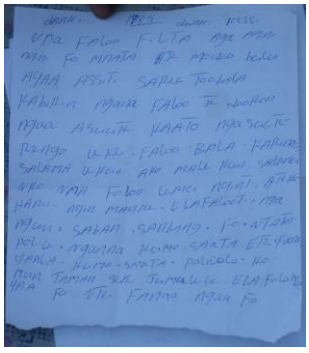
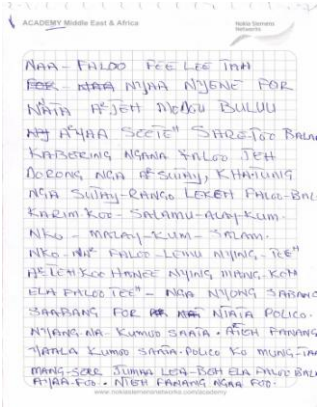
In order to investigate the possible effects of Burama's and Dembo's literacy-backgrounds on their recognition of units in Mandinka, we first compared the spelling as produced by Burama and Dembo.

With respect to this comparison, there are a number of general remarks to be made. First, there are only nine words that are spelled the same in the first two versions, and only seven that are spelled the same in all three versions: *faloo* (occurring twice)*, nko, ntata, saata, yaala, ela*. Note that Burama's original version is a 'natural spelling' and that Dembo's versions are

'exegetic spellings' – spellings produced for the researcher to help him understand Burama's version better. The fact that only a very small number of words are unaltered in Dembo's respelling of Burama's text should perhaps be seen as a consequence of this. Between Dembo's two versions, approximately one out of two words are spelled differently.

Secondly, we also notice a striking difference in capitalisation between the versions. Burama's version uses lower and upper case letters in an unsystematic, somewhat miscellaneous way (e.g., *Nna FAloo FIlITA*) while Dembo's respellings are entirely in capital letters (e.g., *NAA – FALOO FEE LEE TAH* and *NAA FAALO FEELE TAH*). Thirdly, in Burama's spelling there is a rather sparse use of punctuation marking, whereas Dembo punctuates rather abundantly in his first version and not at all in his second version. Fourthly, an innovative feature that is unique to Dembo's first version is the use of superscript as spelling device. This happens for the double (geminated) n in $n^{n}ata$, the aspirated word-final vowels in $seete^{h}$ and tee^{h} and the lengthened /a:/ vowel in $a^{r}jeh$, $a^{r}yaa$, $a^{r}sutay$, na^{r} and a^{r}-teh-koo. This innovation disappears with his second typewritten/e-mailed version, when the medium used to transfer the text does not easily allow for superscript. Table 1 provides an overview of the general differences between the three texts.

Table 1. The three versions of the donkey story

Burama's version	**Dembo's first version**	**Dembo's second version**
		Hi malang How you doing,i am not impress with the position you took despite it was a big maraton NAA FALOO FEELE TAH,,,,,,i lost my donkey NGA NYENEE FOR NINTATA A JEH MODOU BULU,,,,,i search for and i found it with modou AYA SEETEH SARETOO BALA KABERING NGA FALOO JEH,,,,,,he tied the donkey on a cart when when i saw it DORONG NGA SUTAY,,,,i recognise it instantly KHATUNG NGA SUTAY RANGO LEKEH FALOO BALA,,,,because i put an identification mark on the donkey KAREM KO ASALAMU ALAY KUM,,,Karem said peace be upon you thats greetings NKO MALAY KUM SALAM,,,,i said peace be upon you too NKO NAFALO LEMU NYING TEE,,,,,i said these donkey is mine ARTEH KO HANEE NYING MANG KEH ELA FALOO TEE,,,,he said theses not your donkey NGA NYONG SABANG SABANG JEH,,,,we argued FOR NTATA POLICE,,,,untill wereach police NGANA KUMOO SAATA ,,,,,i gave my statement ARTEH FANANG YAALA KUMOO SATA,,,he also gave his statement POLICO KO MUNG TAMANGSERR JUMA LEBEH ELA FALOO BALA,,,the police asked as to wether you put an identification mark on your donkey AYAA FOR NNTEH FANANG NGA FOR,,,,He explained and i explained too END In brief it was a missing donkey that was found in modous position of which he claimed ownership and it spark arguement of which they cannot settle and they finally have to go to the police and the police ask as to wether any of them did put an identity on the donkey and they all explained
handwritten 1985, obtained in July 2008	handwritten in July 2008	type-written, e-mailed on 10th May 2009
title: "dANKI <u>1985</u> danKI KESSI"	no title	no title
miscellaneous use of lower and upper case lettering	entire text in capital letters	entire text in capital letters

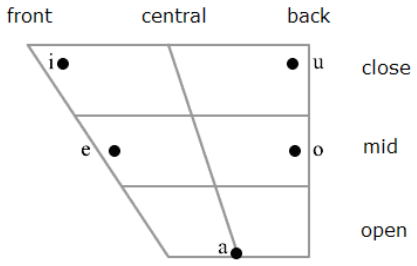
sparse use of punctuation/ hyphenation symbols	abundant punctuation/ hyphenation at both word and sentence-level	no punctuation/hyphenation
plain script only	superscript as spelling device	plain script only

In order to see how grassroots literacy and the learning histories of the informants exactly affected their spelling practices and their ideas on the representation of sounds in letters, a comparative analysis of the informants' use of specific letters for specific phonemes was made against the background of Mandinka's consonant and vowel inventories. In Mandinka, 18 consonants can be distinguished, with 15 places of articulation (see Table 2). In addition to these phonemes there are a number of double consonants or digraphs, i.e. /mb, mf, mp, nj, nd, nt, nc, nk, ns, ng, ny/, and 3 lengthened consonants /ll, mm, nn/. Five vowel phonemes, /i, e, a, o, u/ can be distinguished, that are either short or long (see Table 3).

Table 2. Phonemic consonant inventory of Mandinka

	bi-labial	labio-dental	labio-velar	alveolar	post-alveolar	palatal	velar	uvular	glottal
plosive	p b			t d				k	
nasal	m			n		ɲ	ŋ		
trill				r					
fricative		F		S					h
affricate					tʃ dʒ				
lat. fricative				L					
approximant			w			j			

Table 3. Phonemic vowel inventory of Mandinka



In the Mandinka language materials prepared by the Peace Corps (Colley, 1995b; a) and WEC International (Lück and Henderson, 1993; WEC International, 2002; 2010 [undated]) there are two special characters that do not occur in English, namely ‹ŋ› for the velar nasal /ŋ/ and ‹ñ› with a tilde for the palatal nasal /ɲ/. Although Burama and Dembo's spellings differ in many respects, as can also be seen in Table 4, they agree on ignoring these special characters of the Mandinka alphabet. Both use the diagraphs ‹ng› and ‹ny› to represent these phonemes. The letter ‹c› is prescribed in the language materials as representing the /tʃ/ sound as in the verb *kacaa* 'to chat', but in both Burama's and Dembo's spelling systems, ‹c› is also used to represent /s/ (*police, polico, policolu,* 'police', and *akoroce*). In Burama's spelling we also find the loanword police with an ‹s› (*poliso, polisolu*). In both of Dembo's respellings it is consistently written with ‹c›.

The vowels in Burama's and Dembo's spelling systems present an even more interesting point of comparison as the diversity of sound-letter correspondences and the divergence of the phonemic principle is greater here. The /i/ sound for instance is represented in three different ways in Burama's spelling and in even four different ways in Dembo's first version. The long /i:/ is represented by ‹i› (*filita* 'got lost', *nyini* 'I search', *kabirin* 'when'), ‹ii› (*hanii* 'no') and ‹e› (*e la* 'your', *ete* 'he') in Burama's system and by ‹ee› (*fee lee ta*), ‹ee^{h}› (*tee^{h}*) and ‹e› (*e la*, *nyene*, *kabering*) in Dembo's system. The short, more centralised [ɪ] is represented in all three spelling systems with ‹i› (e.g., *kabering*, *nying*).

The /e/ vowel has three realisations in Burama's and four in Dembo's spellings. The long /e/ is represented by ‹e› (*suute* 'recognise', *nte* 'I', *je* 'see') or by ‹ee› (*saree*, *jee*) in Burama's spelling and by ‹eh› (*nteh*, *jeh*) or ‹ay› (*sutay*) in Dembo's spelling. The short /e/ is represented in Burama's system with ‹e› (*suute-rengo* 'recognition mark') and in Dembo's system with ‹a› (*sutay-rango*) or ‹e› (*lemu* 'is').

The /a/ vowel is represented in two ways in Burama's spelling and in as many as four different ways in Dembo's spelling system. Burama uses ‹a› for the short /a/ (*faloo*, 'donkey', *filita* 'got lost', *bala* 'on') and ‹aa› for the long /a:/ (*ngaa* 'my'). Dembo also uses ‹aa› for the long /a:/ (*naa* 'I', *jumaa* 'where'), but three different graphemes for the short /a/: ‹a› (*faloo* 'donkey'), ‹ah› (*fee lee tah* 'got lost') and ‹a^{r}› (*a^{r}jeh* 'say', *a^{r}sutay* 'recognise').

The /u/ and /o/ vowels present less variety in Burama's and Dembo's spelling systems. The three versions show ‹o› and ‹u› for the short variants /o/ and /u/ and ‹oo› and ‹uu› for the long

vowels /o:/ and /u:/. However, they do not always agree on whether a vowel in a word should be a long or short one (e.g., compare *dooron,* 'always' and *bulu*, 'hand' in Burama's spelling with *dorong* and *buluu* in Dembo's first respelling and *doorong* and *bulu* in Dembo's second respelling).

As can be seen from all these examples, Burama's and Dembo's spelling practices differ in many ways. The sociolinguistic explanation for this difference between Burama's and Dembo's spelling is that they both live in an environment that is poor in terms of access to formal prescriptive rules and support for literacy production in Mandinka. Both are, however, exposed to a fair amount of written English in their private and public lives (e.g., through their children's education, in the linguistic landscape, on television, in newspapers), and Dembo had also enjoyed some literacy education in English. As a result, the normative vacuum for writing in Mandinka seems to be filled with norms that are actively instructed for writing in English, as is especially visible in Dembo's choice for the transcription of vowel phonemes.

Table 4: Word spellings in the three versions

Burama's original version	**Dembo's first respelling**	**Dembo's second respelling**	**Meaning in English**
Nna	NAA	NAA	'my'
FAloo	FALOO	FALOO	'donkey'
FIlITA	FEE LEE TAH	FEELE TAH	'got lost'
Aga	NYAA	NGA	'I'
AyInI			'search it'
nyInI	NYENE	NYENEE	'I search'
FO	FOR	FOR	'until'
MnATA	N^{N}ATA	NNTATA	'came'
AJe	A^{R}JEH	A JEH	'saw it'
Modoo	MODOU	MODOU	'Modou'
bulu	BULUU	BULU	'hand'
AyAA	A^{R}YAA	AYA	'he… it'
ASSITI	$SEETE^{H}$	SEETEH	'tied'
SAREe	SARE	SARE	'cart'
TOO bAlA	TOO BALAA	TOO BALA	'on'
KAbIRIn	KABERING	KABERING	'when'
ngana	NGANA	NGA	'I... my' ; 'I'
FAloo	FALOO	FALOO	'donkey'
Je	JEH	JEH	'see'
dooRon	DORONG	DORONG	'always'
ngaa	NGA	NGA	'I'
ASuute	A^{R}SUTAY	SUTAY	'recognise'
KAATO	KHATUNG	KHATUNG	'because'
nga	NGA	NGA	'I'

suuTe-Rengo	SUTAY-RANGO	SUTAY RANGO	‘recognition mark’
le Ke	LEKEH	LEEKEH	‘put’
FAloo	FALOO	FALOO	‘donkey’
BAlA	BALA	BALA	‘on’
KARInne	KARIM	KAREM	‘then’; ‘Karim’
	KOO	KO	‘say’
SAlAMA le Kun	SALAMU-ALAY-KUM	ASALAMU ALAY KUM	(Arabic greeting)
AKo	NKO	NKO	‘I say’
MAle Kun SAlaMu	MALAY-KUM-SALAM	MALAY KUM SALAM	(Arabic greeting)
NKo	NKO	NKO	‘I say’
Nna	NA^{R}	NA	‘my’
Foloo	FALOO	FALO	‘donkey’
le Mu	LEMU	LEMU	‘is’
nyIn	NYING	NYING	‘this’
TI	TEE^{H}	TEE	‘not’
ATeKo	A^{R}-THE-KOO	ARTEH KO	‘he says’
HAnII	HANEE	HANEE	‘no’
nyIn	NYING	NYING	‘this’
MAnKe… TI	MANG-KEH… TEE^{H}	MANG KEH... TEE	‘is not’
E lA	ELA	ELA	‘your’
FAloo	FALOO	FALOO	‘donkey’
nga	NGA	NGA	‘we’
nyuu	NYONG	NYONG	‘argued and
SAbAn SAnbAng	SABANG SAABANG	SABANG SABANG	argued’
		JEH	‘see’
Fo	FOR	FOR	‘until’
NTATA	NTATA	NTATA	‘we went’
POlICe	POLICO	POLICE	‘the police’
Nganna	NYANG-NA-	NGANA	‘I... my’
KUMO SAaTA	KUMOO SAATA	KUMOO SAATA	‘tell story’
Ete	A^{R}TEH	ARTEH	‘he’
fana	FANANG	FANANG	‘too’
YAAlA	YAALA	YAALA	‘did… his’
KUMO-SAatA	KUMOO SANTA	KUMOO SATA	‘tell story’
POlIColo	POLICO	POLICO	‘the police’
KO	KO	KO	‘say’
Mun	MUNG	MUNG	‘which’
TAMAn sere	TAA MANG-SERR	TAMANGSERR	‘mark’
Juma	JUMAA	JUMA	‘where’
le be	LEHBEH	LEBEH	‘is’
ElA	ELA	ELA	‘you... your’
Falo	FALOO	FALOO	‘donkey’
ba la	BALA	BALA	‘on’
YAA Fo	AYAA FOO	AYAA FOR	‘he said’
NTe	NTEH	NNTEH	‘I’
FAnaa	FANANG	FANANG	‘too’
ngaa	NGAA	NGA	‘I’

F0	FOO	FOR	‘say’

4.3 A comparison of segmentation practices

After the first analysis of the phonological representation of letters as shown in Burama’s and Dembo’s spelling, their boundary-marking and segmentation practices in the sentences that they had provided were investigated, as to see whether their different backgrounds affected their ideas on word-like units in language as well.

In comparing the segmentation practices of the three texts, the first thing that should be noted is that it was hard to define word boundary marking in the handwritings of Burama and Dembo. Previous segmentation studies have focused on disjunction or conjunction of parts of speech, mainly in children’s language segmentation, but in the analysis of the texts we obtained, this distinction was not so obvious. Obviously, the places where clear spaces were used could be counted as boundary markings, but in cases where spaces were small or where hyphens or dots were used, it was unclear whether boundaries were intended to be marked or not. Therefore, segmentation seemed to be less a matter of binary distinctions than of gradual, subtle distinctions. Hence, we propose a continuum from obvious disjunctions in language with use of spaces as boundary markings to obvious conjunctions of language, in which case spaces are not inserted. The use of hyphens and/or dots, which does not show a clear conjunction or disjunction of parts of language, then falls in between.

It can be questioned whether the boundaries in the three versions are boundaries between words, or whether they should be regarded as other boundaries. Since Burama and Dembo’s boundaries mostly correspond to meaningful units that are larger than single morphemes and syllables but smaller than phrases, and since no other regular patterns could be found in the segmentations, we submit that the boundary markings after and between sequences of letters indicate semantic, syntactic or phonological “breaks” at specific points.

Since there were very few items that were segmented in the exact same way in all three spellings, there were numerous items available for comparison. From the first three lines of the donkey story these were *Nna FAloo* (‘my donkey’) for Burama, *NAA – FALOO* in Dembo’s first and *NAA FALOO* in his second respelling. Burama’s *FIlITA* (‘got lost’) occurred as *FEE LEE TAH* in Dembo’s first spelling, and as *FEELE TAH* in his second draft. Similarily, Burama segmented ‘on the cart’ as *SAREe TOObAlA*, whereas Dembo resegmented this first as *SARE-TOO BALAA* and then as *SARETOO BALA*. As is evident from these examples, there were many differences in the places

and ways in which segmentations were made across the texts. Dembo used a lot of hyphens in the first draft of his story, from which we cannot conclude whether he intended to insert a boundary or not, whereas Burama was apparently more extreme in his choice for boundary markings. Tables 4, 5 and 6 place the segmentations of twenty items in the three versions on our three-point continuum. In Table 4, the spelling, grammatical gloss and meaning as given in Colley's (1995a) dictionary are also provided.

Table 4. Segmentations in Burama's text

Line	Joint	Hyphenated	Spaced	dictionary spelling, gloss and meaning
2			NNa FAloo	*naa faloo* (poss N) 'my donkey'
2	FIlITA			*fili –ta* (V-past) 'got lost'
4	ASSITI			*a siti* (pr.3sg V) 'tied it'
4	TOobAlA			*too bala* (postp postp) 'on'
7	le Ke			*le ke* (emph V) 'put'
7			FAloo . BAlA	*faloo bala* (N postp) 'on the donkey'
8			SAlAMA le Kun	*salaa maaleekum*: Islamic greeting
9			le Mu	*le mu* (emph aff) 'is'
9	nyInTI,			*ñin ti* (dem compl)
9	ATeKO			*ate ko* (pr.3sg V) 'he said'
10	MAnKe			*man ke* (neg V) 'is not'
10			E lA	*ila* (poss2sg) 'your'
10	FAlooTI			*faloo ti* (N NEG) 'not… donkey'
11			SAbAn ,SAnbAng	*saba saba* (V V) 'argue and argued'
12	Nganna			*na naa* (N poss1sg) 'I… my…'
13		KUMO-SAatA		*kumoo saata* (N V) 'tell a story'
14			MUn TAMAn sere	*mun taamanseero* (int N) 'which mark?'
14			le be	*le be* (emph V) 'is'
14			ba la	*bala* (postp) 'on'
15			YAA FO	*a ye a foo* (pr.3sg past pr.3sg V) 'he said (it)'
total	9	1	10	

Table 5. Segmentations in Dembo's first respelling

line	Joint	Hyphenated	Spaced
28		NAA – FALOO	
28			FEE LEE TAH
31			…A SEETEH
31			…TOO BALA
34	LEKEH		
34		FALOO-BALA	
35		SALAMU-ALAY-KUM	
37	LEMU		
37		NYING-TEE^{H}	
38		A^{R}-THE-KOO	
38		MAN-KE	
39	ELA		
39			FALOO TEE^{H}
39-40			SABANG SAABANG
41		NYANG-NA	
42			KUMOO SANTA
42-43		MUNG-TAMANG-SERR	
43		LEH-BEH	
43	BALA		
44		AYAA-FOO-.	
total	4	10	6

Table 6. Segmentations in Dembo's second respelling

line	Joint	Hyphenated	Spaced
45			NAA FALOO
45			FEELE TAH
47			…A SEETEH
47			…TOO BALA
50	LEEKEH		
50			FALOO BALA

52			ASALAMU ALAY KUM
54	LEMU		
54			NYING TEE
55			ARTEH KO
55			MANG KEH
55	ELA		
55			FALOO TEE
56			SABANG SABANG
58	NGANA		
59			KUMOO SAATA
60			MUNG TAMANGSERR
60	LEBEH		
60	BALA		
62			AYAA FOR
total	6	0	14

As can be seen from Tables 4 to 6, Burama used joint forms in 9 out of the 20 items that are compared here and inserts spaces in 10 forms, with only one hyphenated form. Dembo used obvious conjunctive forms in only 4 of the 20 respelled items, and inserted spaces in 6 cases. Dembo used hyphens in 10 of the 20 cases in his first hand-written version. Dembo's second, typewritten version differs from his first version in the absence of hyphenation and punctuation. What he had first written with a hyphen is in the second version often not written together at all, but separated by spaces (e.g., SALAMU-ALAY-KUM vs. ASALAMU ALAY KUM). This suggests that the hyphens in his first version indicate disjunction rather than conjunction.

When looking at the type of items that are segmented differently by Burama and Dembo from a semantic and syntactic perspective, we note that differences in boundary marking occur most often on function items, i.e. grammatical forms that mostly lack semantic meaning. Dembo tends to join function items like emphatic markers and possessive markers, while Burama tends to space such items. Instances of these items are: *lekeh* (emphatic marker), *bala* ('on') and *ela* ('your') , both written as one orthographical word – as a sequence of letters marked by spaces – In the hand-written version as well as in the digital version provided by Dembo, but as two

words by Burama. Apparently, Burama sees separate parts within these function items, which are disregarded by Dembo. The idea of not separating these parts, would correspond to the idea that the functors mentioned here only refer to one concept (Van Dyken and Kutsch Lojenga, 1993: 7), and that the parts that Burama distinguishes in these words cannot occur on their own in different places in sentences, or be substituted by other parts of language (Van Dyken and Kutsch Lojenga, 1993: 9-10). This seems indeed to be the case: the separate parts that Burama distinguishes are not found back in other locations, or other surroundings in the three Mandinka texts we obtained.

For tense markers, however, an opposite pattern is visible. In cases where the past tense is marked in the texts (with *yaa* or *-ta*), Burama conjuncts these items to the open class word in front of it, whereas Dembo regards these items as separate words as is evident from the spaces inserted before them. The word *filita* ('got lost') in Burama's text, for instance, is respelled as *fee lee tah* by Dembo. This would suggest that Burama regards the past tense marker as part of the word (as a bound morpheme), whereas Dembo regards it as a free morpheme, in contrast with his treatment of other function items. Apparently, tense markers have a special status for Dembo among the function items: while he does connect function markers usually to other items, just as in English, this is not the case for tense markers.

In this respect, it is worth mentioning though that Dembo's choice for boundary marking mostly coincides with the segmentation of words in English. This is salient in Dembo's use of the form *ela*, which means 'yours', and his writing of *faloo ti* as two separate parts. This separation shows that Dembo is aware of the meaning of the constituent parts of this phrase, namely 'donkey …not'. The separation of *ngan* and *na* by a hyphen in his handwritten version also shows that Dembo somehow distinguishes two elements of form and meaning in Burama's compound *nganna* ('I… my…'). Whether this is coincidental or not, and whether we have to relate this to Dembo's fluency in English or to semantics in general is debatable, but fact is that Dembo shows more segmentation of this kind than Burama.

What we can conclude from this is that it is in the first place not always obvious or easy to define from naturalistic handwritten texts where boundaries are inserted. Even if we see differences in boundary marking between Burama and Dembo in their texts, the continuum on

which they have marked boundaries makes it hard for us to interpret their intentions. It is not clear, for instance, if hyphens are meant to separate or join words, although Dembo's typewritten version suggests that his hyphens were meant as disjunctive rather than conjunctive forms.

Secondly, we noticed that the means with which and the reason for which a text is written may affect the language segmentations made. The use of a keyboard and the medium of e-mail changes Dembo's spelling system. Even if he is rather consistent in replacing the hyphens which he had used in his hand-written version by spaces in the digital version, the fact that he also sometimes writes hyphenated words together in the typewritten version suggests some fluidity in his ideas about units and word boundaries. Thus although Dembo has advanced literacy skills in English, his segmentation of Mandinka is not very stable: boundaries and units are not fully pre-defined. This may be taken as evidence that words do not bear psychological reality, but that they are taught to a considerable extent. Part of the variation between his Dembo's two versions may also be explained by the fact that his spellings are not natural but exegetic spellings directed at the researcher, who requested for additional translation. Dembo's second spelling was made with both Burama's original and his own first respelling available to him. He may have reconsidered his first respelling in an attempt to make the text even clearer for the researcher whom he knew well, and whom he regarded as a foreign friend whom he could help with his study into Mandinka, and changed his segmentations accordingly.

This was also reflected in Burama's text, in which we noticed that Burama's Mandinka literacy education did not lead to consistent boundary marking either. The fact that Burama's spelling differs from the dictionary spelling suggests that even if Burama enjoyed non-formal adult literacy education, boundary marking rules seem to be handled in a fairly loose manner, as was already visible from this study with only one of his texts in which variations in boundary markings occurred. Burama's segmentations seem rather *ad hoc*, not strictly based on formal distinctions in word-types.

For both Burama and for Dembo, word boundaries in Mandinka texts are not fixed, and they do not seem to be based on a standard set of rules or criteria that can be used for boundary marking, as for instance provided by Van Dyken & Kutsch Lojenga (1993). The marking of

boundaries in written Mandinka does not seem to be an automated process for the informants, and the boundaries are not natural or psychologically real, but they are flexible and variable with the writer and text: Burama and Dembo differ in their intuition regarding boundary placement decisions which suggests that those places are not naturally derived from the language itself but are indeed nurtured and probably an outcome of their different sociolinguistic biographies and educational histories as well – as was assumed in previous studies into the relation between unit-recognition in language and literacy.

5 Conclusions

From the comparative analysis of the three versions of the donkey story, we conclude that spelling and segmentation practices not only depend on literacy or education (the 'literacy effect' and the 'schooling effect'), on the language in which one learns to read and write (the 'language effect') or on the writing system (the ‘graphic relativity’, Bugarski, 1993), but also on the type of schooling and the state of literacy (Spolsky, 2009) in the language in question – i.e., on the orthographic regime in society. Our insights from cognitive linguistics added a cognitive lens to the analyses of the ethnographic data in this study, and the ethnographic data showed that merely analysing language segmentation or working with data samples, as is often done in cognitive linguistics, does not account for how people have acquired their literacy skills.

Burama and Dembo both live in an environment that is poor in terms of access to and support for literacy production in Mandinka, or indeed any other local language. As a result, spelling in local languages remains an affair of creativity rather than convention (Kress, 2000). Spellers are left in a normative vacuum, leaving them to spell without orthography. Even though an orthography for Mandinka has been developed, its use is not promoted or enforced in formal education, and not practically ‘enregistered’. Both Burama and Dembo spell and segment in relative freedom, unhindered by institutionalised orthographic rules.

The difference between Burama and Dembo lies in their different learning histories, their educational biographies, and their differently valued routes to literacy. Burama has learned to write Mandinka in adult education and Dembo has learned to write English at school. Burama has not received formal education and has thus not learned to write conventional English. As an

educated man, Dembo never attended adult literacy classes and did thus not learn to write ‘conventional’ Mandinka. As a result of these different personal histories of learning, Dembo’s spelling draws extensively, almost exclusively, on typically English sound-letter correspondences (e.g., <ee> for /i/ and <ay> for /e/), and his segmentation of units in writing also reflects a certain influence from English.

Apparently, differences in background, and the lack of a single set of rules that is commonly known and adhered to, makes Burama and Dembo spell and segment differently: they spell and segment the same items differently within the same text, and they also spell and segment differently on different occasions. The three versions of the donkey story show that segmentation in Mandinka differs with every entextualisation. Word boundaries or units in writing are not fixed in actual Mandinka textual practices. Spelling and segmenting Mandinka is more a matter of creativity than a matter of applying rules. Spelling Mandinka, in sum, is not orthographic, but heterographic.

From this study, it seemed that unit awareness is indeed flexible and not universal, as could be suggested from the work by Bugarski (1993) and Olson (1994). Unit awareness in spoken language has found to be dependent upon the writing system with which one is known, but more than that, it is also based on a writer’s personal educational history, and also on the medium of writing (handwriting vs. digital writing). Scribner and Cole (1981) similarly suggested that one’s personal educational history affects one’s approach to language and writing. The relation between personal educational histories and spelling and the recognition, or marking, of specific units in writing as we have found, suggests that writers develop their ideas on spelling and units in language in their formal, non-formal and informal learning trajectories.

In conclusion, it is important to remember that most of the world’s languages are like Mandinka: they do not have educationally supported and politically endorsed orthographies; they exist in multilingual ecologies in which other languages assume more powerful social positions and fulfill literacy-related roles in society, including formal education. This research therefore begs the question how much of our scholarly knowledge about metalinguistic awareness and the ideas of basic units in language is ‘first-world’ biased. Research on less

studied languages such as Mandinka, in both experimental and ethnographic linguistic research, may contribute to a fuller understanding of the units of language and of writing.

Note

This paper is the product of a dialogue between both authors about their respective work and the research traditions and epistemologies they subscribe to – an ethnographic, sociolinguistic and discourse analytical tradition for Kasper and an experimental and quantitative, cognitive and psycholinguistic tradition for Dorina. Our dialogue was productive in the sense that it allowed us to embark on a journey that brought us to the Sorbonne in Paris (September 2010, 7th International Workshop on Writing Systems and Literacy) and taught us a lot about the idiosyncrasies of our respective research traditions and our own accepted ways of working. The dialogue was not sufficiently productive, however, in the sense that we failed, at least so far, to reconcile and combine both traditions in a way that meets all the criteria and quality requirements for publication in an A-listed journal.

Three earlier versions of this paper were submitted to and judged by the guest editors and two anonymous reviewers of Written Language and Literacy for a special issue following the Paris workshop on “units of language, units of writing”. After elaborate feedback and revision, however, our paper did not stand the test of peer-review and the competition to be published in the special issue. Despite their final decision not to publish the paper, we are grateful to Terry Joyce and Dave Roberts as guest editors for their encouragement, support and feedback along the way. With this acknowledgement, we make this twice-revised document available as a working paper in the Tilburg Papers in Culture Studies series.

References

Agha, Asif. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15: 38-59.

Arnon, Inbal. 2010. *Starting Big: The Role of Multi-word Phrases in Language Learning and Use*. PhD thesis: Stanford University.

Bassetti, Benedetta. 2005. Effects of writing systems on second language awareness: Word awareness in English learners of Chinese as a foreign language. In Vivian Cook and Benedetta Bassetti (eds.) *Second Language Writing Systems*. Clevedon: Multilingual Matters. 335-56.

Basso, Keith H. 1974. The ethnography of writing. In Richard Bauman and Joel Sherzer (eds.) *Explorations in the Ethnography of Speaking*. Cambridge: Cambridge University Press. 425-32.

Baynham, Mike. 2004. Ethnographies of literacy: Introduction. *Language and Education* 18: 285-90.

Berthoud-Papandropoulou, Ioanna. 1978. An experimental study of children's ideas about language. In A. Sinclair, R.J. Jarvella and W.J.M. Levelt (eds.) *The Child's Conception of Language*. Berlin: Springer. 55-64.

Blommaert, Jan. 2008. *Grassroots Literacy: Writing, Identity and Voice in Central Africa*. London: Routledge.

Brandt, Deborah. 1997. The sponsors of literacy. *College Composition and Communication* 49: 165-85.

Bugarski, Ranko. 1993. Graphic relativity and linguistic constructs. In Robert J. Scholes (ed.) *Literacy and Language Analysis*. Hillsdale: Lawrence Erlbaum.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Colley, Ebrima S.Y. 1995a. *Mandinka-English Dictionary*. Banjul: Peace Corps The Gambia.

Colley, Ebrima S.Y. 1995b. *Mandinka Grammar Manual*. Banjul: Peace Corps The Gambia.

Diagne, Pathé. 1978. Transcription and harmonization of African languages in Senegal. In *African languages: Proceedings of the meeting of experts on the transcription and harmonization of African languages*. Paris and Niamey: UNESCO.

Dong, Jie. 2010. The enregisterment of Putonghua in practice. *Language & Communication* 30: 265-75.

DoSE. 2006. *Education Sector Strategic Plan 2006-2015*. Banjul: Department of State for Education.

Fabian, Johannes. 2001. *Anthropology with an Attitude: Critical Essays*. Stanford: Stanford University Press.

Faye, J.C., and M.A. Sillah. 1956. *The Orthography of Gambian Languages: Mandinka and Wolof*. Bathurst: Government Printer.

GBoS. 2007. *The Gambia Multiple Indicator Cluster Survey: Monitoring the Situation of Children and Women. 2005/2006 Report.* Kanifing: Gambia Bureau of Statistics, in collaboration with UNICEF and The World Bank.

Goody, Jack, and Ian Watt. 1963. The consequences of literacy. *Comparative Studies in Society and History* 5: 304-45.

Gorter, Durk. 2006. Introduction: The study of the linguistic landscape as a new approach to multilingualism. In Durk Gorter (ed.) *Linguistic Landscape: A New Approach to Multilingualism*. Clevedon: Multilingual Matters. 1-6.

Homer, Bruce D. 2000. *Literacy and Metalinguistic awareness: A Cross-Cultural Study*. PhD thesis: University of Toronto.

Hoosain, Rumjahn. 1992. Psychological reality of the word in Chinese. *Advances in Psychology* 90: 111-30.

Hughes, Arnold, and David Perfect. 2006. *A Political History of The Gambia, 1816-1994*. Rochester: University of Rochester Press.

Hymes, Dell H. 1974. Studying the interaction of language and social life. In *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.

Juffermans, Kasper. 2010. Local Languaging: Literacy Products and Practices in Gambian Society. Tilburg University.
Juffermans, Kasper, and Caroline McGlynn. 2009. A sociolinguistic profile of The Gambia. *Sociolinguistic Studies* 3: 129-55.
Karmiloff-Smith, Annette, et al. 1996. Rethinking metalinguistic awareness: Representing and accessing knowledge about what counts as a word. *Cognition* 58: 197-219.
Kraak, Albert. 2006. *Homo loquens en homo scribens: over natuur en cultuur bij de taal*. Amsterdam: Amsterdam University Press.
Kress, Gunther. 2000. *Early Spelling: Between Convention and Creativity*. London: Routledge.
Kurvers, Jeanne. 2002. *Met ongeletterde ogen. Kennis van taal en schrift van analfabeten*. Amsterdam: Aksant.
Kurvers, Jeanne, and Helene Uri. 2006. Metalexical awarenees: development, methodology or written language? *Journal of Psycholinguistic Research* 35: 353-67.
Lück, Marlies, and Linda Henderson. 1993. *Gambian Mandinka: Learning Manual for English-Speakers*. Banjul: WEC International.
Lüpke, Friederike. 2011. Orthography development. In Peter Austin and Julia Sallabank (eds.) *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.
Morais, José, et al. 1986. Literacy training and speech segmentation. *Cognition* 24: 45-64.
Olson, David R. 1994. *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge: Cambridge University Press.
Pinker, Steven. 2007. *The Language Instinct: How the Mind Creates Language*. New York: HarperCollins.
Ramachandra, Vijayachandra, and Prathibha Karanth. 2007. The role of literacy in the conceptualization of words: data from Kannada-speaking children and non-literate adults. *Reading and Writing* 20: 173-99.
Sapir, Edward. 1970 [1921]. *Language: An Introductin to the Study of Speech*. London: Hard-Davis.
Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge: Harvard University Press.
Sidibe, B.K. 1979. *A New Orthography of Gambian Mandinka*. Banjul: National Literacy Advisory Committee.
Spolsky, Bernard. 2009. Prolegomena to a sociolinguistic theory of public signage. In Elana Shohamy and Durk Gorter (eds.) *Linguistic Landscape: Expanding the Scenery*. London: Routledge. 25-39.
Tomasello, Michael. 2006. Acquiring linguistic constructions. In William Damon, et al. (eds.) *Handbook of Child Psychology, Volume 2, Cognition, Perception, and Language*. New York: Wiley.
Van Dyken, Julia R., and Constance Kutsch Lojenga. 1993. Word boundaries: Key factors in orthography development. In Rhonda L. Hartell (ed.) *Alphabets of Africa*. Dakar: UNESCO and SIL. 3-20.
Veldhuis, D., and Ad Backus. 2012. More than frequency: Multiword combinations in pre-literate children?s language segmentation. In Saskia Kersten et al. (eds.) *Language learning and language use: Applied linguistics approaches.* Duisburg: Universitätsverlag Rhein-Ruhr. 33-46.
Veldhuis, Dorina, and Jeanne Kurvers. 2012. Offline segmentation and online language processing units: the influence of literacy. *Written language & literacy 15:2,* 165-184.
Veldhuis, Dorina, Jinling Li, and Jeanne Kurvers. 2010. Segmentatie van gesproken Nederlands en Chinees: De invloed van geletterdheid en schriftsoort. *Toegepaste Taalwetenschap in Artikelen* 83: 79-89.
WEC International. 2002. *Mandinka Learning Manual: Short Version*. Banjul: WEC International.
WEC International. 2010 [undated]. *How to write Mandinka: Instructions for Those Who are Used to Write in English*. Banjul: WEC International, Mandinka Literacy Department, www.mandinka.org.