

Lexical Density in Oral versus Written Rangi Texts*

Oliver Stegen

SIL International, oliver@ling.ed.ac.uk

1. Introduction

Comparing oral and written texts in Bantu languages has not been the subject of much study, if only because most Bantu languages have not yet been widely used in writing. With the advent of more and more vernacular literacy programs in sub-Saharan Africa however, opportunities for comparative discourse studies between these two text types increase. As one result of such studies, it will now be possible to investigate whether differences between oral and written texts found in other languages also hold for Bantu languages.

For example, one distinctive feature defined by Halliday (1985) is lexical density (LD), i.e. “the number of lexical items as a proportion of the number of running words” (Halliday 1985:64). “Written language displays a much higher ratio of lexical items to total running words” than spoken language, is one of the claims of Halliday (1985:61). The difference which Halliday draws between LD over the whole text and LD per clause will be referred to below. LD is still used as a defining measure between oral versus written texts, e.g. in Laurén (2002). It is important to keep in mind that the parameter of LD might be indicative more of colloquial versus literary style than merely of oral versus written medium.

In a recent comparative investigation, Stegen (2005: 76) found that for the Tanzanian Rangi language (Bantu F.33), the oral versions of two narratives had a higher LD (56% and 54.7%) than the equivalent written versions (50.3% and 46.6%).¹ The following possible reasons were given for these percentages: that Bantu languages are agglutinating; that the Rangi writer had not yet gained a sufficient level of writing experience; that writers have to compensate certain oral features, e.g. in participant reference.

In this paper, the investigation of LD in Rangi texts is based on a much wider textual basis. In particular, the investigation is extended to other factors potentially influencing LD, e.g. how accomplished a writer/narrator is. Also, LD measurements in Rangi are compared with LD measurements in other languages: two Bantu languages (Swahili and Fuliiru), and an agglutinating language of Peru (Asheninca). Then, some of the discourse features which may be responsible for LD differences between oral and written texts are discussed. Finally, the conclusion summarizes the

* First and foremost, I am deeply indebted to all Rangi writers and narrators who allowed me to make use of their texts, especially Rashid Jumbe, Josephat Lubuva, Andrew Lujuo, Yovin Maingu, Hasani Issa Modu, Peter Patrick and Mama Peter. Thanks go to Peter Patrick and Elizabeth Alphonse Njae for help with typing the texts. For making the Fuliiru texts available to me, I am grateful to Roger Van Otterloo, and for the Swahili texts to Daniel Foster and George Payton. My involvement with the Rangi language project was made possible under residence permit number C 21903 with support from the Anglican Dioceses of Central Tanganyika and of Kondoa, both of whom I thank very much. Financial support from SIL International and Wycliff Germany is gratefully acknowledged. Finally, for feedback on prior drafts of this paper as well as support of various kinds, I thank Brunhilde Bossow and the members of the “Heimat- und Kulturverein Gellersen”, Michael Cahill, Henk Courtz, Helen Eaton, Margaret Hartung, Stephen Levinsohn, Miriam Meyerhoff, Steve Nicolle, Dorothea Stegen, and Hugh Trappes-Lomax. Of course, all remaining errors and misconceptions despite so much help remain entirely my own.

¹ All classifications of Bantu languages in this paper follow Maho (2003).

findings concerning the two questions: how does a Bantu language fare with regard to LD, and what might be contributing factors to differing LD.

2. LD measurements in the 2005 study of Rangi

The previous study (Stegen 2005) was based on only one narrator and one writer (with all texts produced in late 2003). Moreover, the written versions constituted only first drafts for later publication. Table 1 repeats the statistical findings of the previous study. Tokens of the following types of parts of speech were counted as lexical items: nouns, main verbs – i.e. no auxiliary verbs and not the verb ‘to be’ – adjectives, and non-quantifying adverbs; this method is comparable to Laurén’s (2002: 66). For this and all subsequent tables, the following columns are displayed: #w (number of words total), #c (number of clauses total), w/c (average number of words per clause), #lex (number of lexemes total), ld/tx (lexical density across the whole text in percent), l/c (average number of lexemes per clause).

Table 1: Written versus Oral LD (Stegen 2005: 76)

	WRITTEN						ORAL					
	#w	#c	w/c	#lex	ld/tx	l/c	#w	#c	w/c	#lex	ld/tx	l/c
PP1	171	42	4.07	86	50.3%	2.05	182	57	3.19	102	56.0%	1.79
PP2	116	25	4.64	54	46.6%	2.16	223	72	3.10	122	54.7%	1.69

As can be seen, it is actually only LD across the whole text which is greater for the oral versions than for the written versions. LD per clause is by contrast on average higher by 0.35 in the written texts. However, both definitions of LD, originating with Halliday 1985, continue to be used indiscriminately in the literature. For example, studies like Laurén (2002) only use LD over text, whereas studies like Keys (1999: 1050) base their LD measurements on l/c, following further development of LD measurement in Halliday & Martin (1993: 76). Of both definitions, LD per text seems to be more informative as it is independent of clause length. LD per clause, by contrast, seems to be less useful: an l/c value of 2.0 would correspond to a high LD in a 3-word clause, and to a low LD in a 6-word one. Consequently, LD per text percentages are used in this paper unless otherwise stated.

Another observation of potential interest is that Rangi LD figures range higher than those reported for European languages. For example, it is reported for Swedish that “oral texts showed a lexical density of below 40% while the density of written texts usually exceeded 40%” (Laurén 2002: 68). This can be attributed to the fact that Rangi, and Bantu languages in general, are agglutinating, as shown in example (1) (reproduced from Stegen 2005:76), “where every Rangi clause’s lexical density is 100% while the lexical density of the respective English free translations lies between 20-50%.”

- (1) A-ri-viik-ire va-ka-rima va-ka-humula, vi-ryo vi-ka-humula.
 1-5-put-PFV 2-CONS-hoe 2-CONS-finish 8-millet 8-CONS-finish
 ‘After she put it [away], they hoed, they finished, and the millet finished [ripening].’

3. LD measurements in new Rangi texts

Since the 2005 study, the Rangi literacy project has reached production stage, and written Rangi texts have been published and thus become accessible for investigation. First, we have a closer look at the writer of the 2005 study who is also the author of five of the seven stories in a Rangi story booklet (SIL 2005). The two stories PP1 and PP2, which were investigated in Stegen (2005), are published in edited form in the story booklet. Table 2 gives the statistics for all five stories.

Table 2: Written LD by the Same Author (PP)

	#w	#c	w/c	#lex	ld/tx	l/c
PP1ed	248	75	3.31	131	52.8%	1.75
PP2ed	261	78	3.35	135	51.7%	1.73
PP3	137	38	3.61	68	49.6%	1.79
PP4 (AK)	134	48	2.79	82	61.2%	1.78
PP5	228	58	3.93	114	50.0%	1.97

When considering the LD figures of those two stories which were edited out of the oral and the written versions of PP1 and PP2, marked PP1ed and PP2ed in the table, the LD of the edited versions (52.8 and 51.7) also lies between the lower LD of the written originals (50.3 and 46.6) and the higher LD of the oral originals (56.0 and 54.6). This follows from the finding that, in editing, the editors combined oral and written features (Stegen 2005:84). Still, with the one exception of PP4, writer PP's stories exhibit relatively low LD figures, comparable to the original written versions. Upon further investigation, PP reported that he wrote four of the stories from scratch, whereas PP4 was originally written by someone else (AK). Consequently, its 10% higher LD could be accounted for by its original writer's different style. Interestingly enough, the LD parameters of AK's story are quite similar to those of the remaining two stories in the story booklet, written by HI. Table 3 repeats the figures for AK and adds the statistics for HI's stories, including a third one by HI, published in the Rangi primer (SIL 2004).

Table 3: Written LD of a Highly Educated Author (HI)

	#w	#c	w/c	#lex	ld/tx	l/c
AK (PP4)	134	48	2.79	82	61.2%	1.78
HI1	455	149	3.05	275	60.4%	1.85
HI2	724	240	3.02	446	61.6%	1.86
HI3	170	51	3.33	103	60.6%	2.02

The overall LD of HI's stories is higher than either that of the written or oral texts of the 2005 study. In contrast to PP, who has a good intuition about his mother tongue but little experience in writing and a comparatively low education, HI is a highly educated, accomplished author in Swahili (e.g. Issa 1967).² All clauses in table 3, however, are as short as the oral versions of PP (3.19 and 3.10), leading to a lower LD value per clause.

Next, we look at another author who wrote stories for the primer, also a highly educated Rangi writer: AL is a mature postgraduate student with publishing

² Unfortunately, AK's level of education is not known.

experience in both Swahili and English. He produced some very short texts as required by the audience of a primer. Table 4 displays the statistics of AL's texts.

Table 4: Written LD in Short Primer Texts (AL)

	#w	#c	w/c	#lex	ld/tx	l/c
AL1	57	15	3.80	29	50.9%	1.93
AL2	46	8	5.75	23	50.0%	2.88
AL3	16	4	4.00	8	50.0%	2.00
AL4	21	6	3.50	14	66.7%	2.33

With the exception of the very short text AL4, the overall LD of AL's texts (50%) is comparable to PP's (~50%) rather than HI's (>60%). In this case, the decreased text length (caused by the intended audience) seems to be the determining factor for LD. The effect of text length on LD seems to be language specific, however. One of Laurén's groups "produced very short essays, which contributes to the increase of lexical density, since new referents are introduced more frequently than in a longer narrative" (Laurén 2002:70). By contrast, in Rangi, short texts seem to require more copulas and explicit pronouns in their introductions (cf. example 5 in chapter 5, "Discussion of some discourse features", below), both of which lower LD. AL4 is the only one of the four stories which doesn't introduce its participants but starts directly with the situational observation that some children put their hands into corrosive oil, resulting in them being taken to hospital. This lack of introduction might explain AL4's higher LD.

So far, we have looked mainly at written texts. Table 5 shows the statistics of two oral narratives, both tales of Rangi heroes from the past, the first and shorter one about the demise of a cannibal (Mũntaira) who had swallowed an entire Rangi village, and the second and much longer one about Muru, the most famous rainmaker in the history of the traditional Rangi valley of Haubi.

Table 5: Oral LD of Two Narrators

	#w	#c	w/c	#lex	ld/tx	l/c
Mũntaira (RJ)	242	68 (1)	3.56	137	56.6%	2.01
Muru (JL)	2273	629 (9)	3.61	1125	49.5%	1.79

As the average clause length of both stories is almost the same, their differences in LD correspond both across the text and per clause. The first story's overall LD is similar to the oral texts' LD in Stegen 2005 (~55%), while the second story's LD resembles more closely that of PP's original written texts (<50%). A reason for this may be found in the narrators' different background: The first story's narrator is a senior village elder who is accustomed to telling stories, while the second story's narrator is a middle-aged town-dweller who was recorded during the rare occasion of telling his children a story.³ How accomplished a narrator is seems to have an influence on LD.

³ Incidentally, it is the seasoned village narrator who exhibits many more Swahili loanwords in his story, yet the narrator from town is trying painstakingly to avoid Swahili loans, resulting in a very slow start with long pauses as well as a number of false starts or incomplete clauses (their number given in brackets in the #c column).

Looking at the second story, we observed intra-textual LD differences. Table 6 gives an analysis of its different sections. In the column of content description, those sections which consist predominantly of dialogue, i.e. direct speech, have been marked accordingly. The other sections are mainly prose descriptions (JL2 has been marked as containing direct speech but it also contains large chunks of prose).

Table 6: Sectional LD in the Long Narrative (JL)

	description of content	#w	#c	w/c	#lex	ld/tx	l/c
JL1	Introduction, and how Muru's nephew is killed	178	42	4.24	89	50.0%	2.12
JL2	Searching and finding Muru's nephew (dir.sp.)	373	95 (2)	3.93	171	45.8%	1.80
JL3	Killers' denial, and Muru's counsel (dir.sp.)	172	53 (2)	3.25	86	50.0%	1.62
JL4	Four years of rain	185	48	3.85	97	52.4%	2.02
JL5	Muru leaves his sister in the cave (dir.sp.)	176	49 (1)	3.59	78	44.3%	1.59
JL6	Four years of drought	257	58 (2)	4.43	116	45.1%	2.00
JL7	Muru returns to his sister (dir.sp.)	142	50	2.84	81	57.0%	1.62
JL8	Muru's reconciliation with the killers (dir.sp.)	480	153 (1)	3.14	242	50.4%	1.58
JL9	The rain ritual and origin of Rangi clans	310	81 (1)	3.83	165	53.2%	2.04

Looking at overall LD percentages, no pattern seems to emerge. However, LD per clause distinguishes between direct speech sections and prose sections (1.58-1.62 for direct speech versus 2.00-2.12 for prose descriptions, with the mixed section JL2 right in between at 1.80). Correspondingly, clauses in direct speech sections tend to be shorter (2.84-3.59, and 3.93 for JL2) than those in prose descriptions (3.83-4.43). This concurs with our observation that in dialogue, each participant's turn often contains only one lexical item whereas in monologue, clauses often contain two or even more.

Comparing an oral text by one person with a written text by another always has to deal with the challenge of different personal backgrounds. Hence it would be ideal to compare oral and written texts by the same person. An opportunity arose when an amateur historian from Haubi, YM, approached me with drawings and Swahili descriptions of fifty traditional Rangi tools and asked me to find him a publisher. A historical and cultural society in Northern Germany agreed and chose twenty of the tools for publication (Heimat- and Kulturverein Gellersen 2006). It was decided to publish the tool descriptions in three languages: Rangi, Swahili, and German. Hence YM recorded his own Rangi descriptions of the drawings and edited the transcription of this oral rendering into written Rangi. The resulting text type is expository, of course, rather than narrative like the previously discussed texts. Table 7 gives the comparative written versus oral statistics for the first five published tools.⁴

⁴ The oral description of the other tools evolved into a dialogue with other Rangi present during the recording, so has not been considered appropriate for inclusion in this investigation.

Table 7: Written versus Oral LD in Descriptive Texts (YM)

	WRITTEN						ORAL					
	#w	#c	w/c	#lex	ld/tx	l/c	#w	#c	w/c	#lex	ld/tx	l/c
YM1	51	11	4.64	23	45.1%	2.09	83	21	3.95	37	44.6%	1.76
YM2	35	7	5.00	19	54.3%	2.71	85	19	4.47	46	54.1%	2.42
YM4	29	9	3.22	19	65.5%	2.11	52	14	3.71	30	57.7%	2.14
YM6	22	4	5.50	12	54.5%	3.00	26	6	4.33	14	53.8%	2.33
YM8	43	9	4.78	22	51.2%	2.44	70	20	3.50	36	51.4%	1.80

With the exception of YM4, the written versions do not exhibit a significantly higher LD than their oral equivalents. Probably, this is because the written Rangi is based on the oral recording. What distinguishes the two mediums, again with the exception of YM4, is that the written texts tend to contain longer clauses and consequently have higher LD per clause. In a second step, the written Rangi explanations of the tools were compared to their explanation in Swahili which YM had written independently of and prior to the Rangi texts. The main reason behind this comparison was a categorical difference between vernaculars and Swahili in Tanzania. Whereas the former are for the most part only spoken, the latter as national and official language is used in all spheres of life; particularly it is the medium of instruction in primary school and hence, almost every Tanzanian who learns to read and write does so first in Swahili. Table 8 displays the respective statistics.

Table 8: Rangi versus Swahili LD in Descriptive Texts (YM)

	WRITTEN RANGI						WRITTEN SWAHILI					
	#w	#c	w/c	#lex	ld/tx	l/c	#w	#c	w/c	#lex	ld/tx	l/c
YM1	51	11	4.64	23	45.1%	2.09	25	7	3.57	17	68.0%	2.43
YM2	35	7	5.00	19	54.3%	2.71	19	4	4.75	12	63.2%	3.00
YM4	29	9	3.22	19	65.5%	2.11	23	6	3.83	14	60.9%	2.33
YM6	22	4	5.50	12	54.5%	3.00	10	2	5.00	8	80.0%	4.00
YM8	43	9	4.78	22	51.2%	2.44	19	3	6.33	11	57.9%	3.67
YM12	28	7	4.00	14	50.0%	2.00	21	7	3.00	14	66.7%	2.00
YM13	19	4	4.75	10	52.6%	2.50	19	5	3.80	13	68.4%	2.60
YM14	38	7	5.43	17	44.7%	2.43	28	8	3.50	16	57.1%	2.00
YM15	20	4	5.00	12	60.0%	3.00	15	3	5.00	9	60.0%	3.00
YM17	20	4	5.00	13	65.0%	3.25	12	4	3.00	10	83.3%	2.50
YM18	27	5	5.40	15	55.6%	3.00	20	5	4.00	12	60.0%	2.40
YM23	18	6	3.00	12	66.7%	2.00	14	4	3.50	9	64.3%	2.25
YM25	41	9	4.56	20	48.8%	2.22	20	7	2.86	14	70.0%	2.00
YM30	17	5	3.40	12	70.6%	2.40	20	6	3.33	13	65.0%	2.17
YM33	33	7	4.71	15	45.5%	2.14	26	6	4.33	14	53.8%	2.33
YM37	18	4	4.50	11	61.1%	2.75	19	3	6.33	11	57.9%	3.67
YM38	19	5	3.80	13	68.4%	2.60	23	5	4.60	14	60.9%	2.80
YM42	17	5	3.40	10	58.8%	2.00	12	3	4.00	8	66.7%	2.67
YM47	65	16	4.06	30	46.2%	1.88	46	17	2.71	28	60.9%	1.65
YM50	32	6	5.33	18	56.3%	3.00	27	8	3.38	18	66.7%	2.25
total	592	134	4.42	317	53.5%	2.37	418	113	3.70	265	63.4%	2.35

There does not seem to be any one-to-one relation which applies to all twenty texts across the board. However, overall LD across all texts is 10% higher in Swahili than in Rangi (Swahili LD ranges from 53.8-83.3%; Rangi LD from 44.7-70.6%; yet, individual examples where Rangi LD is higher than Swahili LD include YM4, YM23, YM30, YM37, and YM38). This can be taken to corroborate the categorical difference between Rangi as colloquial medium versus Swahili as literary medium.

Looking at LD per clause, the average is the same in Swahili and in Rangi. The main reason for this seems to be that Rangi clauses contain on average more words (17-41 words in 4-9 clauses for the 18 shorter texts, excepting YM1 and 47) than their equivalent Swahili clauses (10-28 words in 2-8 clauses).

4. Comparison with other languages

For two other Bantu languages, we were able to obtain texts in an oral versus a written version each. In Swahili (Bantu G.42d), a personal experience story was recorded during a workshop series in Southern Tanzania, training local literacy workers and translators in discourse analysis and translation principles. The narrator later edited his oral story into a written form. Table 9 displays the statistics of both versions.⁵

Table 9: Oral versus Written LD in a Swahili Text

	#w	#c	w/c	#lex	ld/tx	l/c
oral	222	69	3.22	133	59.9%	1.93
written	192	64	3.00	127	66.1%	1.98

The overall LD of the written version is higher than that of the oral version, and even a bit higher than that of the written Swahili texts of table 8 (63.4%). However, LD per clause is the same, as the oral version contains approximately 15% more words than the written version. On the whole, this Swahili story seems to confirm Halliday's claim for LD per text.

It is not possible to talk about Swahili narrative style or differences between oral and written Swahili, without referring to research done at SOAS, particularly by Maw (1974; 1992). In investigating the questions "To what extent are the same techniques used for story-telling in writing as in speech? or: Are there some characteristics of spoken language as opposed to written that override considerations of intent?" (Maw 1974: 2), Maw compared oral and written Swahili texts, paired with regard to their producers' intention, including reasoned argument, rhetoric persuasion, story-telling, and conversation. While she found the written texts on average to have longer sentences, less clause linkage and more subordination than the oral texts, she found the text intention to have a greater effect on textual differences, i.e. oral and written texts of the same intention showed more similarity to each other than to oral or written texts of a different intention. This may indicate that parameters of style and genre play a greater role than the oral/written divide.

In her 1992 monograph on Swahili oral narrative, Joan Maw makes an interesting observation on one-word clauses which is relevant to LD:

⁵ Only the first three paragraphs, constituting about half of the story, were taken into account as the written version was not completed.

Linguistically, the teller of this story comments throughout with one-word interjections. Sometimes these mark the end of a section, when he might say *Basi* 'That's that', or *Ee* 'Right'. This would seem like the equivalent of paragraphing in a written text. (Maw 1992: 12)

Similar interjections with the same paragraphing function are found in Rangi, e.g. *Baasi*, which literally means 'enough' but is often used to conclude a section. As these words are not counted as lexical items, they lower the overall LD of oral texts. Written texts use punctuation and paragraphing in their stead.

Looking now at a second Bantu language, in a description of Fuliiru (DJ.63) (Van Otterloo & Van Otterloo, forthcoming), the same story, about the hare as thief in the peanut field, is given twice: once written in a formal style, and once transcribed from an oral recording in an informal style, albeit by two different Fuliiru speakers. Van Otterloo identifies some of the distinguishing features between the two styles, e.g. use of the quote marker *ti*, and occurrence of the sequential subjunctive. Table 10 displays the statistics of both versions.

Table 10: Oral versus Written LD in a Fuliiru Text

	#w	#c	w/c	#lex	ld/tx	l/c
oral	303	89	3.40	166	54.8%	1.87
written	361	113	3.19	219	60.7%	1.94

Again, overall LD is higher for the written version, while LD per clause is very similar. This is the case despite the written story being longer than the oral one, and despite the oral version exhibiting slightly longer clauses. On the whole, this Fuliiru example also seems to corroborate Halliday's claim that written texts are lexically denser than oral ones, if basing that claim on LD per text. However, as it also involves a stylistic difference, it is not clear whether the LD difference is due to the formal/informal parameter or due to the oral/written one. In order to differentiate these parameters, it would be necessary to investigate informally written and formal oral texts as well.

We are now turning to a non-Bantu agglutinating language: Asheninca Pajonalino of Peru. Heitzman (1992) investigates its information structure, and particularly redundancy, without being explicitly concerned with LD. Heitzman uses a very different methodology in that she assigns information value, measured in BITS (following Gleason 1970), to each morpheme: 1 BIT to morphemes containing completely new information, 0 BIT to previously given or predictable information, and ½ BIT to morphemes paraphrasing previous ones. Even though measuring something different from LD, namely information load, this categorization is reminiscent of Halliday's suggestion:

[A] list can be drawn up of high-frequency lexical items to be given half of the value of the others. This is equivalent to recognising three categories rather than two: grammatical items, high-frequency lexical items, and low-frequency lexical items. (Halliday 1985:65)

Clauses, words and lexical items have been counted for a short story (Heitzman 1992:50-52), and the results displayed in table 11.

Table 11: LD in a Short Asheninca Story

	#w	#c	w/c	#lex	ld/tx	l/c
Asheninca	31	15	2.07	25	80.6%	1.67

While the overall LD is comparable to that of a couple of short Swahili texts in table 8, it seems to be the case that in Asheninca, many more words in a text contain a lexical root. On the other hand, the root ‘go’ appears so often (6 times) that it is suggestive of grammaticalization. In future LD measurement studies of Bantu texts, grammaticalization, repetition and frequency of lexical items may have to be taken into account.

5. Discussion of some discourse features

It will not be possible to do a comprehensive analysis of discourse features here. However, a number of features had been identified as being used differently in oral versus written Rangi texts (Stegen 2005:82), among them opening and closing devices, the use of conjunctions, and subordinate perfective. In the following, the implications for LD of these and some other features will be discussed.

Opening and closing devices have been mentioned as contributing to lowered overall LD, mainly non-lexical items functioning as paragraphing devices in oral texts (see the discussion on Maw’s (1992) analysis of Swahili narrative style above). These function predominantly in oral or colloquial texts.

Similarly, the frequent use of connectives like *maa*, *dee*, *reeru* or *haaha* is a mark of narration in Rangi. For example, JL (see table 6 above), especially when searching for words, was prone to insert such connectives. In example (2), two clauses with hesitating speech are displayed (the second one is actually incomplete).

- (2) a. Maa ikava reeru haaha kweerire,
and then 9-CONSEC-be obviously now 17-dawn-PFV
‘And then it dawned, ...’
- b. haaha reeru maa dee akafyuka kei labda na,
now obviously and then later 3sg-CONSEC-return again maybe with
‘now later, she then returned again, maybe with ...’

By contrast, the repetition of a verb in perfective form, used as cohesive device, is found in both oral and written texts, and is considered good literary style. Example (3a) is from RJ (oral), and (3b) from PP (written).

- (3) a. Maa akatungya nyika mooto.
and then 3sg-CONSEC-pierce-CAUS grassland fire
Atungirye nyika mooto,
3sg-pierce-PFV-CAUS grassland fire
‘Then he put fire to the grassland. When he had put fire to the grassland, ...’

- b. ... akumbirira iwe isí ya wari.
 3sg-CONSEC-form-APPL 5-stone 9-down-LOC 9-of 14-mush
 Úmbiriire ...
 3sg-form-PFV-APPL
 ‘... then she put a stone inside the mush. When she had put [it inside], ...’

The same can be said of consecutive verb forms used as temporal adverbials, as examples in (4) from different texts show, namely (4a) from JL (oral) and (4b) from AL (written).

- (4) a. Haaha reeru vakapate mulomo uwo,
 now obviously 3pl-CONSEC-receive-SUBJ 3-mouth 3-DEM
 ‘Now, when they received this message, ...’
 b. ... vatwaalwe na sibitáarii. Keende
 3pl-bring-PASS-SUBJ with hospital-LOC since
 takafike sibitáarii,
 1pl-CONSEC-arrive-SUBJ hospital-LOC
 ‘... that they be brought to hospital. Since we arrived at the hospital, ...’

Both these verb forms, the perfective and the consecutive, when used cohesively, cause repetition of lexical items which consequently increases LD.

A distinguishing feature of Rangi in general (as compared to Swahili, for instance) seems to be the frequent use of the copula and forms of the verb ‘to be’ as well as of demonstratives and other pronouns, as exemplified in example (5) from YM12 (with lexical items underlined). While these have a lowering effect on LD, it could not be demonstrated whether they occur more frequently in oral or written texts.

- (5) uhu muti ni ura uri uteréere kamwi wavijáa
 3-DEM 3-tree COP 3-DEM 3-be 3-soft 12-one 3-be-HAB
 ‘... this tree [which they used to carve milking jugs out of] is that absolutely soft one.’

What definitely does increase written LD over against oral LD is the reduction of oral repetition. Example 6 gives evidence of that from the Swahili oral (6a) versus written (6b) story.

- (6) a. nikanunua hivi na vile na vile na vile
 1sg-CONSEC-buy DEM-7 and 7-DEM and 7-DEM and 7-DEM
 ‘And I bought this and that and that and that.’
 b. nikanunua vitu kadhaa vya nyumbani
 1sg-CONSEC-buy 7-thing some 7-of home-LOC
 ‘And I bought some household items.’

This brief overview of a few examples suggests that LD increasing and LD lowering features may be linked to literary/written and colloquial/oral styles respectively. It would be helpful if a future study drew up a more comprehensive list of such features and investigated their occurrence in both oral and written texts, and also in both colloquial and literary styles.

6. Conclusion

The 2005 study had given the following possible reasons for the potentially counter-universal percentages between oral and written texts: that Bantu languages are agglutinating; that the Rangi writer had not yet gained a sufficient level of writing experience; that writers have to compensate certain oral features, e.g. in participant reference. Of these reasons, the agglutinating factor has been discarded above as not contributing to LD differences between the oral and the written medium. However, Bantu languages in general can be expected to have higher LD than languages which are not agglutinating. As corroborating evidence, high LD figures were also attested in Asheninca, a Peruvian agglutinating language.

With regard to LD differences between oral and written texts, it has been shown that a writer's or narrator's experience plays a role. For example, the LD figures of the relatively inexperienced narrator JL and the relatively inexperienced writer PP are similar (around 50% LD), and so are the figures of the experienced narrators MP and RJ and the experienced writer HI (55-60% LD). This seems to have been the overriding factor in the 2005 study, rather than surmised differences in handling participant reference. The wider textual basis did not show evidence of such participant reference differences. However, what does play a role seems to be text length and intended audience as the comparison between texts by the equally experienced writers HI and AL showed.

On the whole, this present study corroborated rather than refuted Halliday's claim of higher LD in written than in oral texts, notwithstanding the ongoing need to investigate the differences between LD per text and LD per clause. It has to be noted, however, that LD is probably more indicative of the colloquial-literary continuum than of the difference between the oral versus the written medium. That is presumably the reason why in the 2005 study, colloquial written texts could have lower LD than oral texts of a supposedly more formalized nature. Moreover, it has been confirmed by Maw's research (1974; 1992) that parameters of style and genre may have a greater effect on textual differences than the oral/written divide.

The following features contribute to lower LD: presence of direct speech, frequent use of opening and closing devices, of connectives, of copula and of pronouns. None of these are necessarily particular to Bantu languages. On the other hand, repetition of verb forms increases overall LD. This seems to be a special feature of Rangi as cohesive device. Whether this is also true of other Bantu languages has to be investigated.

For future research, it will be helpful to control more rigorously for colloquial versus literary style separately from oral versus written medium. In agglutinating languages, it may be necessary to count morphemes rather than words and lexical items. Also, lexical items of different frequency should be assigned different values. If applied together with a list of features which influence LD, such a combined approach should be better able to determine differences between the oral and the written as well as between different styles. Such research is especially relevant in languages which only now start to be used in writing, among them still the majority of the Bantu languages.

References

- Gleason, Henry A. (1970) *Introducción a la lingüística descriptiva*. Madrid: Editorial Gredos. (Translation of *An introduction to descriptive linguistics*. (1961) New York: Holt, Rinehart & Winston.)
- Halliday, Michael A.K. (1985) *Spoken and Written Language*. Geelong, Vic.: Deakin University Press.
- Halliday, Michael A.K. & Jerry Martin (1993) *Writing Science: Literacy and Discursive Power*. Pittsburgh: University of Pittsburgh Press.
- Heimat- & Kulturverein Gellersen (2006) *Alte und neue Heimat eines Kirchgellersers – Mazingira ya Warangi na Wajerumani wa Kale*. Gellerser Hefte, Sonderheft 5. Kirchgellersen: Heimat- & Kulturverein.
- Heitzman, Allene. (1992) Algunas aspectos de la proporción de la información y la redundancia en el ashéninka pajonalino. *Revista Latinoamericana de Estudios Etnolingüísticos* VII: 47-62.
- Issa, Hasani. (1967) *Simba, Ng'ombe na Jogoo na Hadithi Nyingine za Kiirangi*. Dar es Salaam: East African Literature Bureau.
- Keys, Carolyn. (1999) Language as an indicator of meaning generation. *Journal of Research in Science Teaching* 36/9: 1044-1061.
- Laurén, Ulla. (2002) Some lexical features of immersion pupils' oral and written narration. *University of Lund Working Papers in Linguistics* 50: 63-78.
- Maho, Jouni. (2003) A classification of the Bantu Languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.) *The Bantu Languages*. London: Routledge, 639-651.
- Maw, Joan. (1992) *Narrative in Swahili*. London: SOAS.
- Maw, Joan. (1974) *Swahili Style*. London: SOAS.
- van Otterloo, Roger & Karen van Otterloo. (Forthcoming) *The Kifuliiru Language*. Dallas: SIL International.
- SIL (2005) *Símo ja Kílaangi*. Dodoma: SIL International.
- SIL (2004) *Tũsome na kwaandĩka Kílaangi*. Dodoma: SIL International.
- Stegen, Oliver. (2005) Editing Rangi narratives: A pilot study in literature production. *Edinburgh Working Papers in Applied Linguistics* 14: 68-98.