

Use of tense and aspect in Swahili texts¹

Arvi Hurskainen
Department of Languages
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi

Abstract

The report is a statistical survey of the use of various tense and aspect forms of verbs in text. The source texts are the Helsinki Corpus of Swahili 2.0 and the DAHE corpus. The size of the former is about 25 million words, and the latter has less than half a million words. Comparison is made in the use of various tense and aspect forms in different text types.

A more detailed analysis is made with eight fiction books of Shaaban Robert for finding out the degree of consistency in using tense and aspect markers.

Key Words: *tense, aspect.*

1 Introduction

The tense and aspect forms of Swahili differ to some extent from the corresponding forms in many other languages. For example, Swahili has three such tenses that correspond to present tense in other languages, such as English or Finnish. The most common one is the form that means the action that is currently taking place. The marker of this tense is *-na-*. The second is the form that is a sort of general present tense, without special reference to time. Its marker is *-a-*. The third one refers to the habituality of the action. Its marker is *-hu-*, and when it is used, no subject marker is present. All these three forms are generally translated with present tense in other languages.

There are two tenses that correspond to past tense in other languages. Their markers are *-li-* and *-ka-*. The latter is a narrative tense. When something is narrated, the first verb has the *-li-* marker, and when the story continues, the following verbs have the *-ka-* marker.

The perfect tense has a variety of markers, the most common being *-me-*. The marker has also alternative forms, such as *-mesh-*, *-meish-*, and *-mekwish-*. All these three markers are derived from the verb *isha*, which means ending. In other words, to the basic marker *-me-* an additional marker meaning ending is added.

The future tense has two markers, *-ta-* and *-taka-*. Both have the same meaning. The difference is in their use. The marker *-taka-* is used exclusively in relative constructions.

There are five conditional markers, *-nge-*, *-nga-*, *-ngeli-*, *-ngali-* and *-ki-*. Basically, the markers *-nge-*, *-nga-* and *-ki-* refer to present time, that is, if the condition is fulfilled

¹ The report is issued under licence CC BY-NC

presently, the action will take place in future. The markers *-ngeli-* and *-ngali-* refer to past time, that is, if the condition were fulfilled, the action would have taken place. In fact, the marker *-nga-* is redundant, and it does not occur in current texts.

The source texts are the Helsinki Corpus of Swahili 2.0 (HCS) and DAHE corpus. The former has about 25 million words, and the latter has less than half a million words. These corpora differ in that HCS contains mostly written text, with the exception of Parliament speeches, and DAHE only speech, transcribed into written form.

2 Source texts

HCS contains three types of texts. The section *Books* contains about 50 books prose text, mostly fiction, but also some scientific text is included. There are also two sections on news texts, termed here as *News-old* and *News-new*. The only thing that separates these two sections is the time of printing. *News-old* contains texts from the period 1988 to 2002. Early texts were typed from printed newspapers, and when texts started to appear in the web, they were extracted in digital form from various sources. *News new* contains texts from the period 2002 to 2013. All these texts were obtained directly from the web. The fourth section of HCS is *Bunge* that contains speeches of Tanzanian Parliament in years 1994 and 2004-2006. *Bunge* can be considered to include spoken text, because texts are transcriptions of speeches given at the Parliament sessions.

DAHE is a collection of transcriptions of conversations with a large number of people from Tanzanian coast and the islands of Zanzibar, Pemba and Mafia. These texts were collected together with the researchers of Dar-es-Salaam University, and local people in the research area. It must be noted that the collection contains also such texts that are not Swahili. Also metadata reduce the size of the true Swahili text. These limitations must be taken into account when comparing the statistics of DAHE with other texts.

3 Present tense

Swahili has three tenses, each of which corresponds to present tense in many languages. Most common is the tense with the marker *-na-*, for example *ni-na-soma* (I read). We see in Table 1 that it is vastly most common present tense in all text types.

First a word on how the statistics should be read. After each tense type, there is a column showing the relative density of the tense in each text type. First the percentage was counted, and the result was multiplied with 100. As a result, we have numbers that are easily comparable with the corresponding numbers in other columns and in other text types.

On the bottom line, there is the sum of the numbers above, except for the percentage columns. In them, the number is the percentage (multiplied with 100), counted on the basis of the numbers on the bottom line. The average was not counted on the basis of the numbers in the column, because it would distort the result.

Table 1. Present tenses *-na-*, *-a-* and *Habitual* in texts.

	PR:na	% x 100	PR:a	% x 100	HABIT	% x 100	Words total
Books	19223	161	5873	49	4230	36	1189275
News- old	234617	207	35707	32	15025	13	11317510
News- new	146355	276	14653	28	8074	15	5300638
Bunge	251608	267	89399	95	6396	7	9412632
DAHE	5896	13	1528	34	150	3	452032
Total	657699	238	147160	53	33875	12	27672087

The use of the tense *-na-* is especially frequent in *News-new* and *Bunge*. The frequency also in *News-old* is quite high. The reason may be that the texts deal with contemporary issues. On the other hand, the frequency in *Books* is lower, which again may be due to text types. In DAHE, the frequency is especially low. This is due to the nature of the texts and the high percentage of such text that is not standard Swahili.

The clearly highest frequency in the use of the tense *-a-* is *Bunge*. This again may be interpreted so that the speeches in the parliament refer often to general statements on matters rather than on events taking place currently. *Books* have higher frequency than both news texts in this tense type.

The breakdown of the use of the *Habitual* tense is different compared with the other two tenses. *Books* dominate clearly, the news texts are in the middle category, and *Bunge* is low. No obvious reasons can be found for the differences. However, one can hint to the possibility that the older texts use the habitual tense more often than newer texts.

When we compare these three present tenses with each other, we see that the *-na-* tense dominates, the *-a-* tense is the second, and the *Habitual* is least frequent. The result is as expected.

4 Perfect tense

The basic marker of the perfect tense is *-me-*. There are also three other variants that add to the basic marker aversion from the verb *kwisha* (to end). The result is one of the following: *ni-mesha-soma*, *ni-meisha-soma*, or *ni-mekwisha-soma*. These forms refer to the absolute finishing of the action.

As the Table 2 shows, the dominating marker is *-me-*. The other forms have varying frequencies. Surprisingly, the marker *-mekwisha-* has the highest frequency in *Books*. In other text types, the marker *-mesha-* has the highest frequency.

We can conclude that the standard marker for expressing the finished action is *-me-*. If there is need to emphasize the ending of the action, then the marker *-mesha-* is selected. It is the shortest one of the three alternatives. The rarest marker is *-meisha-*. In it, the verb *-isha-* is without infinitive prefix, while in *-kwisha-* it is present.

Table 2. Perfect tenses -me-, -mesha-, meisha, and -mekwisha- in texts.

	PERF: me	% x 100	PERF: mesha	% x 100	PERF: meisha	% x 100	PERF: mekwisha	% x 100	Words total
Books	10377	87	140	1	11	0	521	4	1189275
News- old	149902	132	2855	3	113	0	692	1	11317510
News- new	59262	63	1047	2	92	0	232	0	5300638
Bunge	98722	105	3955	4	303	0	1165	1	9412632
DAHE	1771	39	68	2	3	0	9	0	452032
Total	320034	116	8065	2,9	522	0,2	2619	1	27672087

5 Future tenses

The future tense marker is -ta- in affirmative and negative sentences. In addition, there is the marker -taka- that is used only in relative expressions, such as *a-taka-ye-ku-j-a* (he who will come).

The statistics in this category are interesting. *Bunge* texts use more often the future tense than other texts. This is perhaps due to the fact, that discussions often concern plans that will take effect in future. On the other hand, *Bunge* texts use -taka- only in occasional cases.

Taking both types of markers together, *Bunge* texts still dominate. In using the -ta- marker, *Books* and both news texts are quite similar.

Table 3. Future tenses -ta- and -taka- in texts.

	FUT:ta	% x 100	FUT:taka	% x 100	FUT-ALL	% x 100	Words total
Books	7314	61	730	6	8044	68	1189275
News- old	68905	61	14641	13	83546	74	11317510
News- new	31846	60	6174	12	38020	72	5300638
Bunge	73062	78	9796	1	82858	88	9412632
DAHE	674	15	43	1	717	16	452032
Total	181801	66	31384	11	213185	77	27672087

6 Past tenses

Swahili has two past tense markers, -li- and -ka-. They are used together so, that when the story starts, the first verb uses -li- marker, and the subsequent verbs use -ka- marker. This is the case particularly when the sentence contains several verbs. When the story continues,

the *-li-* marker may be used again. This happens especially when the story moves to another phase in the event. However, there are no clear rules in their use. An experienced storyteller knows when to use this or that marker.

Books has the highest frequency in both categories. Especially the score of the Narrative form is high compared with other text types. This may be due to text types. Fiction books often narrate long stretches of text, where the *-ka-* marker is appropriate.

Table 4. Past tenses *-li-* and *-ka-* in texts.

	PAST	% x 100	NARR:ka	% x 100	Words total
Books	56308	473	11265	95	1189275
News old	401796	355	33631	3	11317510
News new	169880	32	13345	25	5300638
Bunge	133780	142	28579	3	9412632
DAHE	2382	52	4977	11	452032
	764146	276	91797	33	27672087

7 Conditional tenses

There are four conditional tenses, with the markers *-ki-*, *-nge-*, *-ngeli-* and *-ngali-*. The *-ki-* marker refers to the condition to be fulfilled, so that something will happen. For example: *Nikipata fedha nitanunua gari* (If/when I get money, I will buy a car). The other three conditional tense markers are used in contexts such as this: *Ningenunua gari nikipata fedha* (I would buy a car if/when I get money). The markers *-ngeli-* and *-ngali-* refer to condition in past time. Example: *Ningalinunua gari ikiwa ningalikuwa na fedha* (I would have bought a car, if I would have had money).

The difference between markers *-ngali-* and *-ngeli-* is vague, and often they seem to be interchangeable. When we look at the scores of these two markers in various texts, we see a clear trend towards using *-ngeli-* instead of *-ngali-*. The *Books* and *DAHE* corpus clearly prefer *-ngali-*, but all other texts prefer *-ngeli-*. In *Bunge* texts, the difference is huge.

Table 5. Conditional markers *-ki-*, *-nge-*, *-ngeli-* and *-ngali-* in texts.

	COND:ki	% x 100	NGE	% x 100	NGELI	% x 100	NGALI	% x 100	Words total
Books	2695	23	1979	17	95	1	255	2	1189275
News old	25336	22	7207	6	456	0	131	0	11317510
News new	12742	24	4646	9	175	0	21	0	5300638
Bunge	6206	7	13896	15	439	0	13	0	9412632
DAHE	66	1	82	2	3	0	6	0	452032
Total	47045	17	27810	10	1168	0,4	426	0,2	27672087

8 The use of tense and aspect markers in Shaaban Roberts' texts

We will take a closer look at how a single writer uses tense and aspect markers in texts. Shaaban Robert is a famous writer, who was among the very first writers, who used Swahili in his fiction books. In HCS we have eight of his books, and we have a possibility to see how he uses various verb forms.

8.1 Conditional forms

The statistics of the use of conditional forms are in Table 6. Also negative conditional markers *-singe-* and *-singali-* have been added to the table.

The first observation is that he does not use at all the marker *-ngeli-*, although it appears on all other text types. According to Table 5, the marker *-ngeli-* is more common than the marker *-ngali-* in such texts as both news texts, and especially in *Bunge*.

The second observation is that the use of markers varies greatly in different publications. In *sh-adi*, *sh-kus*, and especially in *sh-was*, the *-ngali-* marker is more common than the *-nge-* marker. In the rest of the books, the *-nge-* marker dominates.

The variation corresponds also to the negative markers. If the *-nge-* marker dominates, also the negative *-singe-* marker dominates in the text.

How should we understand this great variation in the use of conditional markers? Is it due to the text types? Or has the mood of the writer influenced the use of markers? Basically, the rule should be that when the verb refers to the time in future, the *-nge-* marker should be used. Correspondingly, if the verb refers to the past time, the marker *-ngali-* should be used. This does not, however, hold in all texts. Even Shaaban Robert used the *-ngali-* marker when referring to the present time or the future.

Later on, in news texts and in *Bunge*, the marker *-ngeli-* seems to have taken the place of *-ngali-*, although the *-nge-* marker still dominates as the general conditional marker.

Table 6. Conditional markers -nge-, -ngali-, -singe- and -singali- in Shaaban Robert's texts.

	nge	% x 100	ngali	% x 100	singe	% x 100	singali	% x 100	Words total
sh-adi	5	3	31	20	0	0	14	9	15191
sh-ins	28	12	11	5	4	2	2	1	23050
sh-kie	67	19	7	2	12	3	1	0	34701
sh-kuf	12	8	9	6	3	2	4	3	15879
sh-kus	14	7	20	12	1	1	7	4	16908
sh-pam	13	10	0	0	1	1	0	0	13168
sh-san	7	6	1	1	0	0	0	0	11709
sh-was	1	1	61	36	0	0	24	14	16831
Total	147	8,25	140	10,25	21	1,13	52	3,88	147437

8.2 Present tense in Shaaban Robert's texts

The statistics in the use of present tense forms in Table 7 deviate greatly from the statistics of other texts. While the most common tense marker is usually *-na-*, followed by *-a-* in density (Table 1 above), here the order is opposite. The *Habitual* tense dominates, and the *-a-* form follows, and the least common is the *-na-* tense. The result is astonishing, because I am not aware of any other text, where this density order can be found.

Another observation is that not all texts have this density order. Especially in texts *sh-ins* and *sh-kie* the *Habitual* tense dominates strongly. In all other texts either *-a-* marker or *-na-* marker dominates.

One reason for this exceptional use of present tenses may be that the author wrote about various general topics of life, which have no clear reference to time. In such texts the *Habitual* form is appropriate.

Table 7. Present tense markers *-na-*, *-a-* and *Habitual* in Shaaban Robert's texts.

	PR:na	% x 100	PR:a	% x 100	Habitual	% x 100	Words total
sh-adi	30	20	28	18	28	18	15191
sh-ins	171	74	157	68	620	269	23050
sh-kie	120	35	358	103	800	231	34701
sh-kuf	42	26	118	74	77	48	15879
sh-kus	100	59	53	31	84	50	16908
sh-pam	175	133	242	184	174	132	13168
sh-san	117	100	164	140	59	50	11709
sh-was	12	7	28	17	5	3	16831
Total	767	56,75	1148	79,38	1847	100,13	147437

8.3 Past tenses in Shaaban Robert's texts

The density of past tense forms varies greatly in Shaaban Robert's texts. In *sh-adi*, every tenth word is a past tense form with the marker *-li-*. Also the frequency of the *-li-* marker is almost ten times more common than the *-ka-* marker. Also in *sh-kus* the *-li-* marker is very frequent.

On the other hand, *sh-ins*, *sh-kie*, *sh-pam* and *sh-san* are texts, where the number of both markers is not very different. However, in all texts the *-li-* marker dominates, which is quite natural, because it is the basic past tense marker.

Table 8. Past tense markers *-li-* and *-ka-* in Shaaban Robert's texts.

	PAST	% x 100	NARR:ka	% x 100	Words total
sh-adi	1668	1098	196	129	15191
sh-ins	163	71	111	48	23050
sh-kie	211	61	150	43	34701
sh-kuf	823	518	50	31	15879
sh-kus	1049	620	71	42	16908
sh-pam	78	59	51	39	13168
sh-san	109	93	63	54	11709
sh-was	1479	879	97	58	16831
Total	5580	424,88	789	55,5	147437

8.3 Perfect tenses in Shaaban Robert's texts

In Shaaban Robert's texts, the perfect tense markers *-mesha-* and *-meisha-* are absent. They are very rare also in other texts. The text *sh-san* differs from other texts in that it uses the *-me-* marker very frequently, and never the marker *-kwisha-*. Also in other texts the marker *-kwisha-* is very rare or absent.

Table 9. Perfect tense markers *-me-* and *-mekwisha-* in Shaaban Robert's texts.

	PERF:me	% x 100	PERF:mekwisha	% x 100	Words total
sh-adi	96	63	0	0	15191
sh-ins	104	45	1	0	23050
sh-kie	151	44	5	1	34701
sh-kuf	117	74	3	2	15879
sh-kus	97	57	6	4	16908
sh-pam	110	84	3	2	13168
sh-san	156	133	0	0	11709
sh-was	53	31	6	4	16831
Total	884	66,38	24	1,63	147437

9 Conclusion

In the survey we see interesting difference in the use of tense and aspect markers in Swahili texts. I have handled the texts in large units, which is likely to hide local differences. However, we see differences that are due to text type and due to the time when the text was written.

The conditional marker *-ngeli-* appears very seldom in older texts, and in Shaaban Robert's texts it never appears. In later texts it is more common that the marker *-ngali-*.

The use of present tense markers also varies greatly. The standard order of frequency is *-na-*, *-a-* and *Habitual*. In Shaaban Robert's texts, which represent the oldest layer of texts. The order is clearly the opposite. The Habitual tense predominates.

A detailed analysis of individual writers might bring to light also other interesting differences.