Chapter 25

A corpus study of the Swahili demonstrative position

Mohamed Mwamzandi

Synchronic studies on Swahili adnominal demonstratives have not addressed the interplay between syntactic position and pragmatic function of these structures. This study shows how referential givenness of discourse entities may explain Swahili word order variation in Swahili adnominal demonstratives. Class 1 (animate nouns) demonstratives are examined in the two attested word orders: NP+DEM and DEM+NP. A close analysis of dataset extracted from the Helsinki Corpus of Swahili reveals that the two structures have distinct pragmatic values. The NP+DEM order is used for active topics while the DEM+NP order reactivates semiactive/inactive topics. This study reveals how the syntax-pragmatics interplay may explain distinct structures viewed as semantic equivalents by native speakers.

1 Introduction

This paper exploresword order variation in Swahili adnominal demonstratives via corpus analysis. The term "adnominal demonstrative" is used in the literature to distinguish demonstratives that co-occur with nouns from stand-alone pronominal demonstratives. While an adnominal demonstrative forms a constituent with an adjacent noun, a pronominal demonstrative is a noun phrase in its own right. More specifically, I analyze the pragmatic use of Swahili demonstratives as outlined by Fillmore (1975; 1982; 1997). Thereafter, I present a qualitative and quantitative analysis of the pre and postnominal position of the Swahili demonstrative. I focus on the relationship that exists between cognitive level of the hearer on discourse entities and the choice of referring expressions (Chafe 1987; Ariel 1988; 1991; 2001; Gundel et al. 1993).

Swahili has various proximal and distal demonstrative forms that obligatorily agree with the nominal class of the noun they modify as exemplified in Table 1.

Notice that the hV- stem is used for the proximal demonstrative while the *-le* stem is used for the distal demonstratives. Further, the agreement affix varies with noun class hence *yu-* and *wa-* for class 1 and 2 and *u-* and *i-* for class 3 and 4.

Besides the semantic distinction of distal and proximal demonstratives, there are two demonstrative constructions that vary in their word order: NP + DEM as seen in (1) and DEM + NP as seen in (2).

Mohamed Mwamzandi. A corpus study of the Swahili demonstrative position. In Jason Kandybowicz, Travis Major, Harold Torrence & Philip T. Duncan (eds.), *African linguistics on the prairie: Selected papers from the 45th Annual Conference on African Linguistics*, 413–429. Berlin: Language Science Press. DOI:10.5281/zenodo.1251756

Table 1: Proximal and distal demonstrative forms of the first four noun classes.

Noun class	Proximal Dem	Distal Dem
1	*hu-yu*	*yu-le*
2	*ha-wa*	*wa-le*
3	*hu-u*	*u-le*
4	*hi-i*	*i-le*

(1) [Msichana yule] a-li-ingia.
1.girl 1.DIST.DEM 1.SM-PST-enter
'That girl entered.'

(2) [Yule msichana] a-li-ingia.
1.DIST.DEM 1.girl 1.SM-PST-enter
'That girl entered.'

The distal demonstrative *yule* 'that' is postnominal in (1) but prenominal in (2). The general tendency in studies on the demonstrative position in Bantu is to claim that the postnominal demonstrative (1) is the unmarked form reserved for the basic gestural function, while the prenominal demonstrative (2) is an innovation aimed at marking definite reference (Ashton 1944; Carstens 1991; 2008; Tamanji 2006). Amidu (2006) points out that both the pre and postnominal demonstrative positions as seen in (1) and (2) can be referential (anaphoric due to previous mention) but does not discuss the pragmatic implications of these demonstratives.

In my analysis, I first discern the adnominal demonstrative function as gestural, anaphoric, or recognitional (Fillmore 1975; 1982; 1997; Himmelmann 1996; Diessel 1999). Thereafter, I qualitatively and quantitatively analyze the pragmatic function of the Swahili demonstrative position. I posit that the postnominal demonstrative as seen in (1) indicates that the intended referent is "active" (Chafe 1987). On the other hand the prenominal demonstrative as seen in (2) indicates that the intended referent is "semi-active" or "inactive". "Semiactive" referents are those discourse entities reintroduced in the discourse after topic shift (change of topic) as well as discourse entities within the conversational context. Topic in this study is what an utterance is about. "Inactive" topics are (re)introduced in the discourse after a long gap of absence or are familiar to the interlocutors.

It is important to note here that Chafe's (1987) activation states as outlined above do not make specific claims on the relationship that exists between activation level and forms of referring expressions. To tackle this absence of matching activation level with forms of referring expression, I invoke the Accessibility Hierarchy (Ariel 1988; 1991; 2001) and the Givenness Hierarchy (Gundel et al. 1993) cognitive theories which associate referential choice with "referential givenness": The awareness level of interlocutors to discourse entities (Gundel & Fretheim 2006). These two cognitive hierarchies rank

demonstrative expressions as mid-accessibility markers. Pronouns are ranked higher than demonstratives while explicit NPs are ranked lower.

A few things on the scope and limitations of this study are worth mentioning. In this study, I examine class 1 (animate nouns) proximal (*hu-yu*) and distal (*yu-le*) demonstratives. Class 1 is chosen because of the relative prominence and sustainability of animate nouns as opposed to inanimate nouns in discourse (Givón 1976; 1983). The applicability of the results is therefore limited to class 1 demonstratives though an extension of the findings to other noun classes is plausible. Further, this study does not look at the distribution of referential demonstratives. Referential demonstratives such as *huyo* are formed by suffixing the "O" of reference to the proximal demonstrative and then deleting the final vowel of the demonstrative (Ashton 1944). While the referential demonstrative is mainly used in discourse to mark definiteness, the use of a proximal/distal demonstrative is not limited to this function (See §2.2.1). Due to its difference in form and functional limitation, the distribution of the referential demonstrative is left out for future research.

The organization of the rest of the paper is as follows. In §2 I explain the methodology. In §3 I present and discuss the results of the study. §4 presents the conclusion and theoretical implications.

2 Methodology

In this section, I explain extraction of the dataset from the Helsinki Corpus of Swahili in §2.1. I then discuss how the dataset was coded in §2.2.

2.1 Extraction of the dataset from the corpus

The source of data in this study is the Helsinki Corpus of Swahili (HCS) which has 14 annotated corpora. The corpora contain current newspaper articles as well as excerpts of literary texts, education and science material written in the mid to late 20th century. Due to the absence of annotations on anaphora resolution in the corpora, I limit the analysis to the HCS books (cf. Mitkov 1994). The HCS books sub-corpus has 1,055,425 words in 71 documents. The documents are mainly Swahili literary texts and education manuscripts.

To obtain the dataset, four queries were made in the HCS. Due to limitations associated with functionality of corpus software, the queries asked for all nouns adjacent to demonstratives whether the demonstrative and the adjacent noun formed a syntactic unit or not. Thus, a manual postediting process aimed at eliminating all the DEM+NP collocations that did not form a syntactic unit was done. Most of these cases were ditransitive verbs with a demonstrative adjacent to both the direct object and the oblique argument as seen in (3).

(3) Njoo u-m-pat-i-e [kijana huyu] [maji ya kunywa].
come 2SG-OM-get-APPL-IMP teenager PROX.DEM water of drinking
'Come and give this teenager some water to drink.'

In (3), the proximal demonstrative *huyu* is modifying the direct object *kijana* 'teenager' but was also displayed by the HCS concordancer as a prenominal demonstrative modifying the indirect object *maji ya* kunywa 'drinking water'. Other cases that were eliminated include pronominal identification demonstratives in which the copula introducing the demonstratum was deleted; and adnominal demonstratives from poems whose pre or postnominal position may be driven by metrical requirements. Table 2 shows the number of adnominal demonstratives before and after disambiguation.

Table 2: Adnominal demonstratives before and after postediting.

And-Dem	And-Dems before disambiguation	And-Dems after disambiguation	Difference
Prenominal proximal	133	109	24
Postnominal proximal	135	124	11
Prenominal distal	140	126	14
Postnominal distal	114	75	39
Total	522	434	88

Each of the disambiguated demonstrative expression was then displayed in its narrow context (in the HCS of Swahili concordancer) as well as its wider context (in the original text) for contextual and statistical analysis.

2.2 Coding the data

Each demonstrative expression was coded for the following variables: dem-type (proximal, distal), dem-function (gestural, anaphoric, recognitional), dem-position (prenominal, postnominal) and the activation state (active, semiactive, inactive). Anaphoric demonstratives were further coded for referential distance. While coding for dem-type and dem-position was straightforward after displaying the queries in their wider context, coding for the dem-function, referential distance and activation state needs further elaboration. Each of these variables is explained in turn in §2.2.1, §2.2.2 and §2.2.3.

2.2.1 Demonstrative function

Adnominal demonstratives as referring expressions have mostly been analyzed by looking at the demonstrative function: gestural, anaphoric, and recognitional. Coding for these demonstrative functions is explained below.

'Gestural' here does not necessarily mean actual pointing but rather situations which need 'pointing' of some sort to establish reference. In the dataset there are instances such as (4) where a cue word may indicate that the demonstrative in question is gestural.

(4) [Yule bwana] u-na-mu-on-a: Mzalamo yule?
DIST.DEM person 2SG-PRS-OM-see-FV Zaramo DIST.DEM
'Do you see that person: is he a Zamoro (ethnic community)?'

In (4), the demonstrative expression *yule bwana* 'that person' was coded gestural because the verb *on-a* 'see' draws the attention of the hearer to a potential discourse entity within the conversational context. Only first mentions of referents within conversational contexts were coded as gestural. Subsequent mentions were coded as anaphoric.

Anaphoric demonstratives track discourse entities across clauses (intra-sentential) (5) as well as across sentences (inter-sentential) (6) (Botley & McEnery 2000).

(5) A-li-po-fik-a kwa [mzee Malongo], [mzee yule]
SM-PST-when-arrive-FV at old.man Malongo old.man DIST.DEM
a-ka-shangaa.
SM-SEQ-surprise
'When he (Kiliilo) arrived at mzee Malongo's home, that old man (Malongo) was surprised (to see him).'

(6) a. U-ki-vuka bahari saba, ku-na [chewa]$_i$ mkubwa.
2SG-COND-cross seas seven, 17SM-be grouper big
'If you cross the seven seas, there is a grouper (type of fish).'

b. [Chewa huyu]$_i$ a-ki-vuta pumzi
1grouper this 1SM-COND-breath air
'When this grouper is breathing…'

In (5), the NP *mzee Malongo* in the matrix clause is the antecedent of the demonstrative expression *mzee yule* 'that mzee' in the embedded clause. In (6a), the noun *chewa* 'grouper' is the antecedent of the demonstrative NP *chewa huyu* 'this grouper' in (6b). Demonstratives used to track referents in intra and intersentential contexts were coded as anaphoric.

Demonstratives used recognitionally indicate common knowledge and therefore do not have a co-specification element in the surrounding situation or preceding discourse (Diessel 1999). This is illustrated in (7).

(7) [Huyu Juma] ka-shindw-a ku-ku-tunz-a.
PROX.DEM Juma SM.PRF-defeat-FV INF-OM-take.care-FV
'This Juma has failed to provide for you.'

In (7) the proximal demonstrative *huyu* indicates that *Juma* is the man the speaker and the hearer all know. The demonstrative expression here is not anaphoric since the referring expression *Juma* has no apparent antecedent in the preceding discourse. It is also not gestural because the referent *Juma* is not physically present in the conversational context.

Although recognitional demonstrative expressions are overwhelmingly used in first mentions to indicate common knowledge, there are instances where subsequent mentions via a demonstrative expression may mark the referent as familiar at that point of discourse. This is illustrated in (8).

(8) Kumbe [yule mtu mweupe] amba-ye a-li-kuwa a-me-nusur-ik-a
INTJ DIST.DEM man white COMP-REL SM-AUX SM-PERF-save-STV-FV
ku-ua-w-a na wenyeji
INF-kill-pass-FV by natives
'Alas, that white man who had escaped being killed by the natives … '

In (8), the referential distance between the adnominal demonstrative and its antecedent was 118 clauses. The writer is aware of the "referential problem" (Auer 1984) caused by topic shift and therefore adds more information to the adnominal demonstrative in form of a restrictive clause to ensure successful identification of the referent. Following Himmelmann (2006: 230), in addition to first mention of discourse entities to indicate common knowledge, I also coded demonstrative expressions as recognitional if the gap of absence after previous mention was too long to warrant "additional anchoring or descriptive information to make the intended referent more accessible".

2.2.2 Referential discourse

Referential distance has been described as the most important diagnostic tool for measuring referential givenness. Givón (1983: 36), for example, explains that the effect of referential givenness on accessibility correlates with other factors such as interference from other possible discourse entities since "a high referential distance would show - all other things being equal - more interfering topics in the preceding register." Interfering topics are other topics mentioned other than the immediate topic before its previous mention in the discourse.

Since the finite clause is the locus for topic update, referential distance in this study is the number of finite clauses from the relevant adnominal demonstrative expression to a co-specifying explicit NP to its left (cf. Kameyama 1998; Poesio et al. 2004; Taboada & Zabala 2008). The clause as the 'locus for topic update' implies that it is at the clause level that more information about the topic is added. Example (9) illustrates how coding for referential distance with the finite clause as the unit of analysis was done. Notice that each of the clauses in (9a-c) contains new information about the topic (*mjumbe* 'messenger'). The letter *u* stands for 'utterance' – the minimal unit of analysis in discourse, in this case, the finite clause (cf. Grosz et al. 1995).

(9) a. (u1) mjumbe wa tano alipotakikana, (u2) alitokea bila ya ajizi. (u3) Mjumbe huyu alikuwa Ridhaa
[Mjumbe] wa tano a-li-po-tak-ik-an-a,
messenger of fifth SM-PST-when-want-STV-RECP-FV
When the call for the fifth messenger was made,

b. a-li-toke-a bila ya ajizi
SM-PST-appear-FV without of fail
he came forth without fail.

c. [Mjumbe huyu] a-li-kuwa Ridhaa.
messenger PROX.DEM SM-PST-AUX Ridhaa
This messenger was Ridhaa.

The ref-distance in (9) was coded as 2, that is, there are two finite clauses before the subsequent mention of the topic *mjumbe* in (9c).

2.2.3 Activation states

Depending on the referential distance between the adnominal demonstrative under consideration and its antecedent, the adnominal demonstrative in question was coded as active, semi-active or inactive. A question that arises under this description adapted from Chafe (1987) is: What is the number of intervening utterances that qualify a discourse entity to be active/semiactive/inactive?

The intended referent of an active referent is within the immediate consciousness of the discourse participants. Thus, an adnominal demonstrative expression was coded as 'active' if there was an apparent antecedent in the preceding utterance as is the case in (10).

(10) a. Mtu wa pili ku-kut-an-a na-ye a-li-kuwa [mzee].
person of second INF-meet-RECP-FV with-3SG SM-PST-AUX old.man
'The second person to meet me was an old man.'

b. [Mzee huyu] a-li-kuwa a-ki-peleka ng'ombe wake mtoni.
old.man PROX.DEM SM-PST-AUX SM-IPFV-take cows his river-LOC
'This old man was taking his cows to the river.'

In (10a) the NP *mzee* 'old man' is an apparent antecedent of the adnominal expression *mzee huyu* 'this old man' in (10b). The adnominal demonstrative expression *mzee huyu* 'this old man' in (10b) was therefore coded as active.

Semiactive referents in this study were of two types: situational (in conversational context) and textual (in discourse texts). Consequently, all gestural adnominal demonstratives were coded as semiactive. In discourse texts, a referent was coded as semiactive if there was an intervening topic(s) between the previous explicit mention of the antecedent NP to the adnominal demonstrative expression under consideration. This is illustrated in (11) and (12).

(11) Mbele ya-ngu ku-li-kuwa bado watu wawili [yule mzee] na
in.front POSS-1SG LOC17-PST-AUX still people two DIST.DEM old.man and
[msichana mmoja].
girl one
'In front of me, there were still two people remaining, that old man and one girl.'

(12) [Msichana huyu], a-li-ye-kuwa bado a-me-weka kitambaa
girl PROX.DEM SM-PST-REL-AUX still SM-PRF-put handkercheif
'This girl, who still had a handkercheif placed … '

In (11), yule mzee 'that old man' and *msichana mmoja* 'a girl' are the potential topics for the following utterance. A potential topic is a referent within an utterance that can be chosen by the speaker to be the center (topic) of the next utterance (cf. Grosz et al. 1995). In the following 4 utterances (not presented above), the *mzee* 'old man' is established and continued as the topic. In (12), the demonstrative expression *msichana huyu* 'this girl' reintroduces the girl mentioned in (11). The adnominal demonstrative *msichana huyu* 'this girl' in (12) was therefore coded as semiactive because of the interfering topic, *mzee* 'old man'.

All recognitional demonstratives were coded as "inactive" because their identification depends on retrieval of the discourse participants from the memory (see §2.2.1).

3 Results and discussion

In this section, I discuss the relevance of the demonstrative function in explaining the demonstrative position in §3.1. I then discuss the relationship that exists between the demonstrative position and activation states (active, semiactive, inactive) in §3.2.

3.1 Demonstrative function and position

Of the 434 adnominal demonstratives in the dataset, gestural demonstratives were 52, anaphoric 308 and recognitional 74. The frequencies of dem-type (proximal and distal) in both the pre and postnominal position are presented in Table 3.

Table 3: Dem-function and dem-position in proximal and distal demonstratives.

	Gestural			Anaphoric			Recognitional		
	Pre	Post	Total	Pre	Post	Total	Pre	Post	Total
Proximal	38	9	47	49	110	159	22	5	27
Distal	2	3	5	83	66	149	41	6	47

The pragmatic value of the demonstrative position for each of the demonstrative functions will be discussed in turn.

3.1.1 Gestural demonstratives

Table 3 above shows that the proximal gestural demonstratives in prenominal position were 38 and 9 in postnominal. There were 2 distal gestural demonstratives in prenominal

position and 3 in postnominal. These frequencies show that, first, the proximal demonstrative is mostly used as the deictic expression for the gestural function. The total frequency of proximal gestural demonstratives is 47 while the total frequency of the distal gestural demonstratives is 5. This frequency difference is significant (X^2 (1,N=52)=33.92, $p < 0.001$). Second, the gestural demonstratives have a higher frequency count in prenominal position than postnominal. The total number of prenominal demonstratives is 40 while in the postnominal position the total number is 12. This frequency difference is also significant (X^2 (1,N=52)=15.08, $p < 0.001$).

The difference in the demonstrative position for the gestural demonstratives can be explained by recalling the grammaticalization of the Swahili prenominal demonstrative to express definite reference (Ashton 1944; Givón 1976; Carstens 1991; 2008). In their paper on definite reference in English, Clark & Marshall (1981: 38), mention PHYSICAL COPRESENCE (presence in conversational contexts) as one of the reasons which license definite reference in English. Based on the contextual analysis of the corpus data, I posit that the prenominal demonstratives are mostly used to point to definite referents due to PHYSICAL COPRESENCE as seen in (4) repeated here as (13).

(13) [Yule bwana] u-na-mu-on-a: Mzalamo yule?
DIST.DEM person 2SG-PRS-OM-see-FV Zaramo DIST.DEM
'Do you see that person: is he a Zamoro (ethnic community)?'

Based on the high frequency of gestural demonstratives in prenominal position, it can be deduced that the prenominal position is mainly used to mark the referents as accessible (semi-active) in conversational contexts. The examples in (14) and (15) further illustrate this.

(14) Huyu kondoo tu-m-pelek-e kwa Mfalme Ndevu.
PROX.DEM sheep 1PL-OM-take-IMP to King Ndevu
'This sheep, let us take her to King Ndevu.'

(15) Mfalme a-ki-m-pat-a kondoo huyu a-ta-furahi sana.
King SM-SBJV-OM-get-FV sheep PROX.DEM SM-FUT-happy very
'If the King gets this sheep, he will be very happy.'

In (14), because the discourse participants are all aware of the sheep's presence within the conversational context, the prenominal demonstrative in *huyu kondoo* 'this sheep' marks the referent as definite due to PHYSICAL COPRESENCE. In (15), however, the postnominal position of the demonstrative marks the previously mentioned *kondoo* 'sheep' as anaphoric. As it will be seen in §3.1.2, the unmarked position for anaphoric demonstratives is postnominal.

3.1.2 Anaphoric demonstratives

The distribution of the 308 anaphoric demonstratives was as follows. There were 49 proximal demonstratives in prenominal position but 110 in postnominal position (X^2

(1,N=159)=23.40, p<0.001). The distal demonstratives were 83 in prenominal position but 66 in postnominal position, p>0.05. These results show that for the anaphoric demonstratives the proximal demonstrative has a higher frequency in postnominal position than in prenominal. When contrasted with the distal postnominal demonstrative, the proximal postnominal demonstrative frequency is also significantly higher (X^2 (1,N=176)=11.00, p<0.001). In the prenominal position, the distal demonstrative has a significantly higher frequency than the proximal demonstrative (X^2 (1,N=132)=8.76, p<0.005).

In order to further explore the frequency tendencies of the anaphoric demonstratives, the referential distance of the anaphoric demonstrative expressions in the dataset was analyzed. The results are presented in §3.1.3.

3.1.3 The effect of referential distance on the anaphoric demonstrative position

In measuring the referential distance, the number of finite clauses between an adnominal demonstrative and a co-referential NP to its left was counted and recorded in a database. The raw data was then log transformed to reduce the skewness of the data distribution. After log-transformation, the Shapiro-Wilk test revealed that the data distribution for the distal and proximal prenominal demonstratives was normal, p>0.05. The skewness of the distal and proximal postnominal demonstrative data was greatly reduced but not completely eliminated, p<0.05.[1]

Table 4 and Table 5 report the descriptive statistics of the demonstrative position for the raw and log-transformed data respectively. The number in parentheses is the standard deviation while the number outside the parentheses is the mean referential distance.

Table 4: Mean referential distance and standard deviation of raw data.

Dem_Type	Prenominal	Postnominal
Proximal	5.55 (5.39)	5.25 (5.06)
Distal	7.40 (6.55)	5.29 (4.35)

Table 5: Mean referential distance and standard deviation of log-transformed data.

Dem_Type	Prenominal	Postnominal
Proximal	1.34 (0.87)	1.25 (0.92)
Distal	1.70 (0.77)	1.30 (0.89)

[1]The statistical operations conducted in this study assume normal distribution. Log transformation of the variables enhances normal distribution, hence reducing the influence of outliers on the results (Baayen 2008).

The mean referential distance of the prenominal demonstratives is higher than that of postnominal demonstratives. A non-repeated measures ANOVA with ref-distance as the dependent variable and dem-type and dem-position as the independent variables reveal that there is a significant main effect of ref-distance on dem-type ($F(1,308)=6.09$, $p<0.05$) and dem-position ($F(1,308)=5.90$, $p<0.05$). There was no significant interaction between dem-type and dem-position, $p>0.05$.

Further, a planned comparison using the t.test reveals that the mean ref-distance of the distal prenominal demonstrative is higher that of the proximal prenominal demonstrative, $p<0.05$. The nonparametric Wilcoxon test applied to compare the median of the distal prenominal demonstrative and the distal postnominal demonstrative indicates that the medians of these two vectors and their distributions are different. Hence, the mean referential distance of the distal prenominal demonstrative is also higher than that of the postnominal distal demonstratives, $p<0.05$. However, there is no significant difference between the mean ref-distance of the proximal pre and postnominal demonstratives. These statistics show that:

1. The difference in referential distance between the proximal and distal postnominal demonstrative is not significant.
2. The distal prenominal demonstrative tends to be separated from its antecedent by longer referential distance than the distal postnominal demonstrative as well as the proximal pre and postnominal demonstrative.
3. The difference in referential distance between the proximal pre and postnominal demonstratives is not significant.

I illustrate these observations with examples from the corpus.

These statistics show that the proximal demonstrative is frequently used in postnominal position when the referential distance is short (See example (15)). The insignificant difference in referential distance between the proximal and distal postnominal demonstratives further suggests that there are cases when a distal postnominal demonstrative may be used after a short referential distance as seen in (16).

(16) a. Adili a-li-po-taka ku-ingia ndani,
Adili SM-PST-when-want INF-enter inside
'When Adili was about to go inside (the house), '

b. a-li-ona [mtu] a-me-simama mlango-ni
SM-PST-see person SM-PRF-stand door-LOC
'he saw a person standing at the door … '

c. Adili a-li-dhani [mtu yule] a-li-kuwa bawabu.
Adili SM-PST-assume person DIST.DEM SM-PST-AUX security.officer
'Adili thought that the person was a security officer.'

The referent *mtu* 'person' introduced in (16b) is continued in (16c). The postnominal position of the demonstrative in (16c) marks the referent as 'active'. The use of the postnominal distal demonstrative *yule* 'that' instead of the proximal demonstrative *huyu* 'this' has a special effect of marking the "narrative distance" (Leonardo 1987; Wilt 1987), that is, the author is narrating events from a third person's perspective. In the third person's perspective style of narration, the narrator is not involved in the events of the story.

Further, the results show that the distal prenominal demonstrative is separated from its antecedent by long referential distance as illustrated in (17).

(17) [Yule msichana] a-li-ingi-a.
DIST.DEM girl SM-PST-enter-FV
'That girl entered.'

In (17), the demonstrative expression *yule msichana* reintroduces the girl as the topic after 45 clauses.

It is important to mention here that most corpus generalizations are based on statistical tendencies (See Mwamzandi 2014 for more examples). In general, anaphoric proximal and distal demonstrative are used postnominally after a short referential distance to mark the intended referent as active. Anaphoric distal demonstrative are used prenominally after topic shift to mark the referent as semiactive.

3.1.4 Recognitional demonstratives

The frequency of recognitional proximal demonstratives in prenominal position was 22, and 5 in postnominal position (X^2 (1,N=27)=10.70, $p < 0.01$). In prenominal position, the frequency of distal demonstratives was 41, and 6 in postnominal position (X^2 (1,N=47)=26.06, $p < 0.001$). The difference between the recognitional demonstratives in pre and postnominal positions is statistically significant (X^2 (1,N=74)=36.54, $p < 0.001$). It can be inferred from the results that a demonstrative is preferred in prenominal position if used recognitionally.

Contrary to Himmelmann's (1996) claim that only one of the demonstratives, mostly the distal demonstrative, is preserved for the recognitional function across languages, both the distal and proximal demonstratives can be used for this function in Swahili as seen in (18) and (19).

(18) [Yule mtoto wako] a-na-ye-fundisha Chuo Kikuu,
DIST.DEM child your SM-PRS-REL-teach university
'That child of yours who teaches at the university, …'

(19) Hii ni kazi ya majirani zetu, hasa [huyu mjukuu wa Ndenda].
9PROX.DEM is 9work of neighbours our especially PROX.DEM grandchild of Ndanda
'This is the work of our neighbours, especially this grandchild of Ndenda.'

In (18), the speaker uses the distal demonstrative *yule* 'that' to signal familiarity. However, in (19) the proximal demonstrative *huyu* 'this' signals not only "larger situation" familiarity (Hawkins 1978) but also "community membership", that is, the referent (mjukuu wa Ndenda) lives within the speaker's neighborhood (Clark & Marshall 1981). The use of the distal demonstrative expression *yule mjukuu wa Ndenda* in (19) instead of the proximal demonstrative *expression huyu mjukuu wa Ndeda* eliminates the "community membership" implication.

3.2 Activation states

In this section, I discuss the effect of the active, semiactive and inactive activation states on the form of the adnominal expression in the following paragraphs in turn.

As mentioned earlier, subsequent mentions of referents via anaphoric demonstrative expressions if the referent was a continued topic were coded as active. Table 6 presents the frequencies of the demonstrative expressions coded as active.

Table 6: Demonstrative expressions coded as active.

	Prenominal	**Postnominal**	**Total**
Proximal	42	88	130
Distal	32	47	79

A few things can be said about these frequencies. First, the frequencies show that the proximal demonstrative is used more frequently than the distal demonstrative if the activation state of the intended referent is active (X^2 (1,N=209)=12.45, $p < 0.001$). Second, there is a higher frequency of proximal demonstrative in postnominal position than in prenominal position if the activation state of the intended referent is active (X^2 (1,N=130)=16.28, $p < 0.001$). Third, though insignificant, the frequency of the distal demonstrative in postnominal position is higher than in prenominal position if the activation state of the intended referent is active, $p > 0.05$. These results corroborate the statistics I presented on the effect of referential distance on demonstrative position of anaphoric demonstratives in §3.1.3.

All gestural demonstratives as well as demonstratives used anaphorically after topic shift were coded as semi-active. Table 7 and Table 8 present the frequencies of the gestural and anaphoric semi-active demonstratives.

Table 7: Semiactive gestural demonstratives.

	Prenominal	**Postnominal**	**Total**
Proximal	38	9	47
Distal	2	3	5

Table 8: Semiactive anaphoric demonstratives.

	Prenominal	Postnominal	Total
Proximal	7	22	29
Distal	51	19	70

I have discussed the significance of the gestural demonstrative frequencies in pre and postnominal position in §3.1.1. Here I discuss the frequencies of the anaphoric demonstratives coded as semiactive. First, the frequencies of the anaphoric semiactive demonstratives show that the distal demonstrative is used more frequently than the proximal demonstrative (X^2 (1,N=99)=16.98, p < 0.001). Second, the frequency of proximal postnominal demonstrative is higher than proximal prenominal demonstratives (X^2 (1,N=29)=7.76, p < 0.01). Third, the frequency of distal demonstrative in prenominal position is higher than postnominal (X^2 (1,N=70)=14.63, p < 0.001).

As for inactive activation state, all first mentions of familiar referents and subsequent mentions of discourse entities after a long referential distance via adnominal demonstrative expressions were coded as inactive. Table 9 presents the frequencies of the adnominal demonstratives coded as inactive.

Table 9: Inactive adnominal demonstratives.

	Prenominal	Postnominal	Total
Proximal	22	5	27
Distal	41	6	47

I have also discussed the significance of these frequencies in §3.1.4. In summary, the prenominal position is used more frequently than postnominal if the referent is inactive.

4 Conclusion and theoretical implications

The observation that pre and postnominal demonstratives can be used as referring expressions for discourse entities has ramifications on the analysis of Swahili demonstrative expressions in pragmatics as well as syntax. In pragmatics, it has been observed cross-linguistically that activation level of topics may be represented via different forms of referring expressions (Gundel et al. 1993; Ariel 1988; 1991; 2001). In Swahili, the demonstratives co-occur with the noun to mark different activation levels of referents. The results of this study show that postnominal demonstratives are high accessibility markers, prenominal demonstratives are mid-accessibility markers, and prenominal demonstratives followed by a restrictive clause are low accessibility markers.

This functional role of the demonstrative position independently motivates a syntactic analysis of the Swahili demonstrative in pre and postnominal position (Carstens 1991;

2008). In the postnominal position, the unmarked order of Swahili noun modifiers is: N>POSS>DEM>Quantifier (20) (cf. Rugemalira 2007).

(20) eneo langu hili lote
area 5AGR.POSS.1SG PROX.DEM all
'all this area of mine'

Of these three types of modifiers, only the demonstrative may occur prenominally. The functional distinction of the demonstrative in pre and postnominal position as observed in this study rules out the possibility of these demonstratives orders being manifestation of a single abstract syntactic structure. The different N + DEM/DEM + N constructions correspond to different discourse needs.

Acknowledgements

I thank Dr. Laurel Stvan, Dr. Jason Kandybowicz and Dr. Jeffrey Witzel for their input on earlier versions of this paper. I also thank the ACAL reviewers for their comments and recommendations.

Abbreviations

Unless indicating person, numbers in glosses indicate noun class. Abbreviations follow Leipzig Glossing Rules, with the following exceptions:

FV	final vowel	SEQ	sequential
INTJ	interjection	SM	subject marker
OM	object marker	STV	stative

References

Amidu, Assibi A. 2006. *Pronouns and pronominalization in Swahili grammar*. Rüdiger Köppe Verlag: Köln.

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24(1). 65–87.

Ariel, Mira. 1991. The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16(5). 443–463.

Ariel, Mira. 2001. Accessibility theory: An overview. In Ted Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*, 29–87. Amsterdam/Philadelphia: John Benjamins.

Ashton, Ethel O. 1944. *Swahili grammar (including intonation)*. London: Longmans, Green & CO LTD.

Auer, Peter. 1984. Referential problems in conversation. *Journal of Pragmatics* 8.5-6. 627–648.

Baayen, Harald R. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press: New York.

Botley, Simon & Mark Anthony McEnery. 2000. *Corpus based and computational approaches to discourse anaphora*. Amsterdam/Philadelphia: John Benjamins.

Carstens, Vicki May. 1991. *The morphology and syntax of determiner phrases in Kiswahili*. UCLA dissertation.

Carstens, Vicki May. 2008. DP in Bantu and Romance. In Cecile De Cat & Demuth Katherine (eds.), *The Bantu-romance connection*, 131–166. Amsterdam/Philadelphia: John Benjamins.

Chafe, Wallace L. 1987. Cognitive constraints on information flow. In Russel S. Tomlin (ed.), *Coherence and grounding in discourse. Typological studies in language*, vol. II, 21–52. Amsterdam: John Benjamins.

Clark, Herbert H. & Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber & Ivan A. Sag (eds.), *Elements of discourse understanding*, 10–64. Cambridge: Cambridge University Press.

Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. John Benjamins: Amsterdam/Philadelphia.

Fillmore, Charles J. 1975. *Santa Cruz lectures on deixis, 1971*. Bloomington: Indiana University Linguistics Club.

Fillmore, Charles J. 1982. Towards a descriptive framework for spatial deixis. In Robert J. Jarvella & Wolfgang Klein (eds.), *Speech, place and action: Studies in deixis and related topics*, 31–59. Chichester/New York/Brisbane/Toronto/Singapore: John Wiley & Sons LTD.

Fillmore, Charles J. 1997. *Lectures on deixis*. CSLI Publications: Stanford, California.

Givón, Talmy. 1976. Topic, pronoun, and grammatical agreement. In Charles N. Li (ed.), *Subject and topic*, 149–188. New York: Academic Press.

Givón, Talmy (ed.). 1983. *Topic and continuity in discourse* (Typological studies in language 3). Amsterdam/Philadelphia: John Benjamins.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.

Gundel, Jeanette K. & Thorstein Fretheim. 2006. Topic and focus. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 175–196. Malden, USA: Blackwell Publishing.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Hawkins, John A. 1978. *Definiteness and indefiniteness*. Groom Helm: London.

Himmelmann, Nikolaus P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In Barbara Fox (ed.), *Studies in anaphora*, 205–254. John Benjamins: Amsterdam/Philadelphia.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mose (eds.), *Essentials of language documentation*, 1–30. Berlin: Mouton de Gruyter.

Kameyama, Megumi. 1998. Intrasentential centering: A case study. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering theory in discourse*, 89–112. New York: Oxford University Press.

Leonardo, Robert. 1987. Response to Wilt: Discourse distances and the Swahili demonstratives. *Studies in African Linguistics* 18(1). 97–105.

Mitkov, Ruslan. 1994. An intergrated model for anaphora resolution. Proceedings of the 15th International Conference on Computational Linguistics. Kyoto, Japan. In.

Mwamzandi, Mohamed. 2014. *Swahili word order choices*. University of Texas at Arlington. Ph.D dissertation.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio & Janet Hitzeman. 2004. Centering: A parametric theory and its instantiations. *Computational Linguistics* 30(3). 309–63.

Rugemalira, Josephat M. 2007. The structure of the Bantu noun phrase. *SOAS Working Papers in Linguistics* 15. 135–148.

Taboada, Maite & Loreley Hadic Zabala. 2008. Deciding on units of analysis within Centering Theory. *Corpus Linguistics & Linguistic Theory* 4(1). 63–108.

Tamanji, Pius N. 2006. *Concord and DP structure in Bafut. Afrikanistik online.*

Wilt, Timothy. 1987. Discourse distances and the Swahili demonstratives. *Studies in African Linguistics* 18(1). 81–95.