Relative Constructions of Swahili and Machine Translation

Arvi Hurskainen
Institute for Asian and African Studies, Box 59
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi

Abstract

The report discusses the problems of translating relative constructions from Swahili into English. Various types of relative constructions are introduced first, and then solutions are introduced for handling them in MT. Emphasis is in optimality in each case. Because the translation system is composed of several sequential modules, it is feasible to test various solutions and evaluate the effects of each solution. The discussion of problems and solutions is illuminated with examples.

Key Words: *machine translation, relative constructions*

Abbreviations:
MT   machine translation
SL   source language
SVO   language type (subject + verb + object)
TL   target language

1 Introduction

One of the major problems in translating Swahili into English is the extensive mismatch between syntactic constituents. This applies particularly to verbs. Whereas English expresses verb-forms and verb phrases using separate words, Swahili often uses bound morphemes, prefixes and suffixes, attached to the verb stem. For MT to succeed, the morphemes in Swahili have to be tagged in such a way that they can be unambiguously manipulated. Such operations include reordering, deletion, substitution and insertion.

The basic word order in verb phrases is SVO (subject-verb-object) in both languages, but there are a number of differences, for instance, in marking the subject and object. Particularly prominent in Swahili is double-marking of the subject. That is, the subject is marked also in the verb, although the sentence has an overt subject. Also the object can be marked twice, in the verb and in the overt object, although there are restrictions to this rule. Because English marks the subject and object only once, the MT translation system must be able to decide when the glosses of the subject and object must be deleted and when retained in the TL.

Relative structures of Swahili constitute another area, where the production of correct translation requires rules with constraint conditions beyond clause boundaries. For example, the presence of the subject marker gloss in the verb of the main clause depends on the presence or absence of the subject that is beyond the relative clause on the left.

Another problem to solve is whether the relative prefix in the verb refers to the subject or to the object. This must be decided, because it affects translation.

Genitive relatives are constructed in Swahili in a way much different from English. Also this needs special treatment.

In rule-based sequential MT, it is important to decide on which point of processing each operation should be performed. For example, should the operations on constituents, such as reordering, substitution, deletion and insertion, be performed before or after the production of surface forms of the target language (TL). Another issue is how to formulate the tag set in the morphological analyzer, so that subsequent rule writing becomes optimal.

In the implementation discussed here, all operations on constituents are performed before the production of surface forms. However, the rules for performing the operations are applied in strict order. Reasons for this are discussed below.

2 Relative structures in Swahili

Relative structures in Swahili can be classified according to their form and function.

2.1 Forms of relative structures

There are three basic methods for forming relative expressions in Swahili.

(1)
 (1a) mtu anayesoma
```
mtu
      "mtu" N 1/2-SG { the } { man }
anayesoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr-OBJ>
```

 (1b) mtu asomaye
```
mtu
      "mtu" N 1/2-SG { the } { man }
asomaye
      "soma" V 1/2-SG3-SP VFIN { he } [soma] { read } SVO **SUBJ-
CLB GEN-REL 1/2-SG { who } @FMAINVtr-OBJ>
```

 (1c) mtu ambaye anasoma
```
mtu
      "mtu" N 1/2-SG { the } { man }
ambaye
      "ambaye" PRON **SUBJ-CLB GEN-REL 1/2-SG { who } @SUBJ
anasoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
```

Among these types, (1c) can be used in all relevant verb forms. Type (1a) can be used with present tense (marker -na-), future tense (marker -taka-), past tense (marker -li-) and present tense negative (marker -si-). Type (1b) is the so-called general relative without reference to time, aspect or mood.

All three types of relative constructions can stand alone (2) or begin an embedded relative clause (3).

(2)

 (2a) Huyu ni mtu anayesoma.

```
*huyu
      "huyu" PRON DEM :hV 1/2-SG { this } CAP @SUBJ
ni
      "ni" V V-BE NOSUBJ { is } @FMAINVintr-def
mtu
      "mtu" N 1/2-SG { the } { man }
anayesoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr-OBJ>
```

 (2b) Huyu ni mtu asomaye.

```
*huyu
      "huyu" PRON DEM :hV 1/2-SG { this } CAP @SUBJ
ni
      "ni" V V-BE NOSUBJ { is } @FMAINVintr-def
mtu
      "mtu" N 1/2-SG { the } { man }
asomaye
      "soma" V 1/2-SG3-SP VFIN { he } [soma] { read } SVO **SUBJ-
CLB GEN-REL 1/2-SG { who } @FMAINVtr-OBJ>
```

 (2c) Huyu ni mtu ambaye anasoma.

```
*huyu
      "huyu" PRON DEM :hV 1/2-SG { this } CAP @SUBJ
ni
      "ni" V V-BE NOSUBJ { is } @FMAINVintr-def
mtu
      "mtu" N 1/2-SG { the } { man }
ambaye
      "ambaye" PRON **SUBJ-CLB GEN-REL 1/2-SG { who } @SUBJ
anasoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
```

(3)

 (3a) Mtu anayesoma sana anafaulu katika maisha.

```
*mtu
      "mtu" N 1/2-SG { the } { man } CAP @SUBJ
anayesoma
```

```
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu"  V  1/2-SG3-SP  VFIN  {  he  }  PR:na  [faulu]  {  be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

 (3b) Mtu asomaye sana anafaulu katika maisha.
```
*mtu
      "mtu" N 1/2-SG { the } { man } CAP @SUBJ
asomaye
      "soma" V 1/2-SG3-SP VFIN { he } [soma] { read } SVO **SUBJ-
CLB GEN-REL 1/2-SG { who } @FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu"  V  1/2-SG3-SP  VFIN  {  he  }  PR:na  [faulu]  {  be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

 (3c) Mtu ambaye anasoma sana anafaulu katika maisha.
```
*mtu
      "mtu" N 1/2-SG { the } { man } CAP @SUBJ
ambaye
      "ambaye" PRON **SUBJ-CLB GEN-REL 1/2-SG { who } @SUBJ
anasoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu"  V  1/2-SG3-SP  VFIN  {  he  }  PR:na  [faulu]  {  be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

Because the subject of the finite verb is always marked in verb, also the constructions in (4) are grammatical.

(4)

(4a) Anayesoma sana anafaulu katika maisha.
```
*anayesoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO CAP @FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu" V  1/2-SG3-SP VFIN { he } PR:na [faulu] { be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

(4b) Asomaye sana anafaulu katika maisha.
```
*asomaye
      "soma" V 1/2-SG3-SP VFIN { he } [soma] { read } SVO **SUBJ-
CLB GEN-REL 1/2-SG { who } CAP @FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu" V  1/2-SG3-SP VFIN { he } PR:na [faulu] { be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

(4c) Ambaye anasoma sana anafaulu katika maisha.
```
*ambaye
      "ambaye" PRON **SUBJ-CLB GEN-REL 1/2-SG { who } CAP @SUBJ
anasoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anafaulu
      "faulu" V  1/2-SG3-SP VFIN { he } PR:na [faulu] { be
successful } SVO AR @FMAINVtr-OBJ>
katika
      "katika" PREP { in }
maisha
      "maisha" N 6-PLSG { the } { life } AR
```

2.2 Functions of relative structures

Relative structures in Swahili can have three types of functions. Below there are examples of each of them.

2.2.1 Relative refers to the subject

 (5a) Mtoto anayempenda mama yake atafaulu.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr+OBJ>
mama
      "mama" N 9/10-SG { the } { mother } AN HUM @OBJ
yake
      "ake" PRON POSS 9/10-SG SG3 { his/her/its } @GCON
atafaulu
      "faulu" V  1/2-SG3-SP VFIN {  he  } FUT:ta [faulu] {  be
successful } SVO AR @FMAINVtr-OBJ>
```

Some cases are ambiguous, such as in (5b). The relative can refer either to the subject or
the object.

(5b) Huyu ni mtoto anayempenda.
```
*huyu
      "huyu" PRON DEM :hV 1/2-SG { this } CAP @SUBJ
ni
      "ni" V V-BE NOSUBJ { is } @FMAINVintr-def
mtoto
      "mtoto" N 1/2-SG { the } { child }
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr-OBJ>
"penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-OBJ-REL { whom }
1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr-OBJ>
```

2.2.2 Relative refers to the object

(6) Mtoto ninayempenda amesafiri.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP
ninayempenda
      "penda" V 1/2-SG1-SP VFIN { *i } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr-OBJ>
amesafiri
      "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

2.2.3 Relative expresses association

(7)
(7a) Mtoto niliye naye amesafiri.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
niliye_naye
```

```
       "niliye_naye" { which *i have }
amesafiri
       "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(7b) Mtoto niliyekuwa naye amesafiri.
```
*mtoto
       "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
niliyekuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST 1/2-SG-SUB-REL { who }
INFMARK AUX-WA SV MONOSLB IDIOM-V> @FMAINVintr
naye
       "naye" <IDIOM { have }
amesafiri
       "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(7c) Mtoto ambaye nilikuwa naye amesafiri.
```
*mtoto
       "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
ambaye
       "ambaye" PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { whom } @SUBJ
nilikuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST INFMARK AUX-WA SV MONOSLB
IDIOM-V> @FMAINVintr
naye
       "naye" <IDIOM { have }
amesafiri
       "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(7d) Mtoto ninayekaa naye amesafiri.
```
*mtoto
       "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
ninayekaa
       "kaa" V 1/2-SG1-SP VFIN { *i } PR:na 1/2-SG-SUB-REL { who }
[kaa] { stay } SVO @FMAINVtr+OBJ>
naye
       "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
       "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(7e) Mtoto ambaye ninakaa naye amesafiri.
```
*mtoto
       "mtoto" N 1/2-SG { the } { child } CAP
ambaye
       "ambaye" PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { whom }
ninakaa
```

```
        "kaa" V 1/2-SG1-SP VFIN { *i } PR:na [kaa] { stay } SVO
@FMAINVtr+OBJ>
naye
        "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
        "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

In (7a), the verbal particle -li- represents the verb 'to be'. When time is expressed for the verb 'to be', the inflecting verb 'kuwa' is used (7b). Te relative expression formed using the pronoun 'amba' can be used in almost all relative expressions (7c). Relative expressions denoting association or possession can be formed also from other verbs (7d-e).

The negative forms of the examples in (7) are given in (8).

(8)
(8a) Mtoto nisiye naye amesafiri.
```
*mtoto
        "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
nisiye
        "nisiye" REL-SI SG1-SP { who am not }
naye
        "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
        "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(8b) Mtoto nisiyekuwa naye amesafiri.
```
*mtoto
        "mtoto" N 1/2-SG { the } { child } CAP
nisiyekuwa
        "wa" V 1/2-SG1-SP VFIN { *i } NEG-REL 1/2-SG-SUB-REL { who }
INFMARK AUX-WA SV MONOSLB IDIOM-V> @FMAINVintr
naye
        "naye" <IDIOM { have }
amesafiri
        "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(8c) Mtoto ambaye sikuwa naye amesafiri.
```
*mtoto
        "mtoto" N 1/2-SG { the } { child } CAP
ambaye
        "ambaye" PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { whom }
sikuwa
        "wa" V NEG SG1-SP VFIN { *i } PAST-NEG:ku AUX-WA SV MONOSLB
IDIOM-V> @FMAINVintr
naye
        "naye" <IDIOM { have }
```

```
amesafiri
      "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(8d) Mtoto nisiyekaa naye amesafiri.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP
nisiyekaa
      "kaa" V 1/2-SG1-SP VFIN { *i } NEG-REL 1/2-SG-SUB-REL { who
} [kaa] { stay } SVO @FMAINVtr+OBJ>
naye
      "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
      "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

(8e) Mtoto ambaye sikai naye amesafiri.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP
ambaye
      "ambaye" PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { whom }
sikai
      "kaa" V NEG-a 1/2-SG1-SP VFIN { *i } [kaa] { stay } SVO
@FMAINVtr+OBJ>
naye
      "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
      "safiri" V 1/2-SG3-SP VFIN { he } PERF:me [safiri] { travel
} SV AR @FMAINVintr
```

The negative forms of the relative constructions in the TL are produced after the operations discussed here, that is, before the actual surface forms of the TL are produced. Therefore, the same rules apply to affirmative and negative forms.

2.2.4 Relative refers to possession

In Swahili, the relative referring to possession or association is constructed following the formula in (9).

(9) Mtoto ambaye mkono wake umevunjika analia.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP @SUBJ+rel
ambaye
      "ambaye" PRON **POS-CLB POS-GEN-REL REL 1/2-SG { whose }
mkono
      "mkono" N 3/4-SG { the } { hand } @SUBJ
wake
      "ake" PRON POSS 3/4-SG SG3 { his/her/its } @GCON
umevunjika
```

```
        "vunjika" V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me [vunja] { get
broken } PREFR SV STAT @FMAINVintr
analia
        "lia" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [lia] { cry } SVO
@FMAINVtr-OBJ>
```

An example of a similar structure with a non-human referent is in (10).

(10) Mti ambao majani yake yamekauka umeanguka.
```
*mti
        "mti" N 3/4-SG { the } { tree } PLACE CAP @SUBJ+rel
ambao
        "ambao" PRON POS-GEN-REL REL 3/4-SG { which }
majani
        "jani" N 5/6-PL { the } { leaf } @SUBJ
yake
        "ake" PRON POSS 5/6-PL SG3 { his/her/its } @GCON
yamekauka
        "kauka" V 5/6-PL-SP VFIN NO-SP-GLOSS PERF:me [kauka] { dry
up } SV EXT: STAT :EXT @FMAINVintr
umeanguka
        "anguka" V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me [angua] { fall
down } PREFR SVO STAT @FMAINVtr-OBJ>
```

The processes needed in translating examples (9) and (10) will be discussed below.

3 Types of operations and their implementation

In manipulating relative structures, basic operations include deletion, insertion , substitution, and constituent reordering. Below are examples of each type of operations.

3.1 Deletion

Because the subject is almost always marked in finite verb constructions[1], also when the overt subject is present, the presence of the gloss of the subject marker in LT must be controlled. Consider the examples (11) and (12).

(11) mtu anayesoma
```
mtu
        "mtu" N 1/2-SG { the } { man } @SUBJ
anayesoma
        "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr-OBJ>
```

(12) anayesoma

---

[1] A notable exception is the non-tense affirmative form with the marker hu-.

```
anayesoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr-OBJ>
```

In (11), the gloss 'he' must be deleted, because an overt subject 'mtu' is present. In (12) the subject prefix is in a pronominal function and the gloss 'he' must be retained.

A similar problem occurs also in object prefixes, especially when the verb has an animate object.[2] Examples are in (13-16).

(13) mtoto anayempenda mama yake
```
mtoto
      "mtoto" N 1/2-SG { the } { child }
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr+OBJ>
mama
      "mama" N 9/10-SG { the } { mother } AN HUM @OBJ
yake
      "ake" PRON POSS 9/10-SG SG3 { his/her/its } @GCON
```

(14) mtoto anayempenda
```
mtoto
      "mtoto" N 1/2-SG { the } { child } @SUBJ
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr-OBJ>
```

(15) anayempenda
```
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr-OBJ>
```

(16) mtoto anayesoma kitabu
```
mtoto
      "mtoto" N 1/2-SG { the } { child } @SUBJ
anayesoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who }
[soma] { read } SVO @FMAINVtr+OBJ>
kitabu
      "kitabu" N 7/8-SG { the } { book } @OBJ
```

In (13), both the subject prefix gloss 'he' and object prefix gloss 'him/her' must be deleted in TL, because the overt subject and object are present. In (14), the subject prefix gloss

---

[2] When the verb has an animate object, it is common to mark the object in the verb also when the verb has an overt object. In the case of non-animate objects the marking is optional, and in most cases it is absent.

must be deleted, but the object prefix gloss retained, because there is no overt object. In (15), both the subject prefix gloss and object prefix gloss must be retained, because there both the overt subject and object is missing. The example in (16) represents the case where the object marker of a non-animate object is missing, and therefore the object marker gloss dies not appear in the verb.

There are two basic methods for writing deletion rules for these cases. In one method, deletion rules are written using the kind of output as seen in above examples. In other words, the subject gloss or object gloss is removed depending on whether an overt subject or object is present. However, this method is rather sensitive to modifications in the output and may lead to tough maintenance problems.

In another method, applied here, each finite verb-form will produce two representations for the subject prefix and object prefix, one with the gloss and another without the gloss. On the basis of this representation it is then possible to write CG rules for selecting or deleting appropriate readings. Consider the example in (17).

```
(17) mtoto anayempenda mama anasoma
mtoto
      "mtoto" N 1/2-SG { the } { child } @SUBJ
anayempenda
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN { he } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO @FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na 1/2-SG-SUB-REL {
who } 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO
@FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na 1/2-SG-SUB-REL {
who } 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO
@FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN { she } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO @FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN { she } PR:na 1/2-SG-SUB-REL { who
} 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO @FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na 1/2-SG-SUB-REL {
who } 1/2-SG3-OBJ OBJ { him/her } [penda] { love } SVO
@FMAINVtr+OBJ>
      "penda" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na 1/2-SG-SUB-REL {
who } 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO
@FMAINVtr+OBJ>
mama
      "mama" N 9/10-SG { the } { mother } AN HUM @OBJ
anasoma
      "soma" V 1/2-SG3-SP VFIN { he } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
      "soma" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [soma] { read }
SVO @FMAINVtr-OBJ>
      "soma" V 1/2-SG3-SP VFIN { she } PR:na [soma] { read } SVO
@FMAINVtr-OBJ>
```

```
       "soma" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [soma] { read }
SVO @FMAINVtr-OBJ>
```

We see that for each subject prefix there is a reading with a gloss, and a reading without a gloss, marked with the tag 'NO-SP-GLOSS'. Similarly, for each object prefix there is a reading with a gloss, and a reading without gloss, marked with the tag 'NO-OBJ-GLOSS'. On the basis of the context, it is possible to write rules for selecting an appropriate reading in each case. After applying the rules, we get the correct readings as shown in (18).

(18) mtoto anayempenda mama anasoma
```
mtoto
       "mtoto" N 1/2-SG { the } { child } @SUBJ
anayempenda
       "penda" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na 1/2-SG-SUB-REL {
who } 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO
@FMAINVtr+OBJ>
mama
       "mama" N 9/10-SG { the } { mother } AN HUM @OBJ
anasoma
       "soma" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [soma] { read }
SVO @FMAINVtr-OBJ>
```

In (18), the reading 'NO-SP-GLOSS' was selected for both verbs. The noun 'mtoto' is the subject of the last verb, and no gloss in the verb is needed. The reason for selecting the reading with 'NO-SP-GLOSS' for the first verb is that because the relative refers to the subject, no gloss in TL is needed. The example (19) shows what happens if the relative refers to the object.

(19) mtoto ninayempenda anasoma
```
mtoto
       "mtoto" N 1/2-SG { the } { child } @SUBJ
ninayempenda
       "penda" V 1/2-SG1-SP VFIN { *i } PR:na 1/2-SG-OBJ-REL { whom
} 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO @FMAINVtr-OBJ>
anasoma
       "soma" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [soma] { read }
SVO @FMAINVtr-OBJ>
```

In (19), the verb with the relative referring to the object retains the subject prefix gloss but deletes the object prefix gloss. Also the verb following the relative clause loses the subject prefix gloss, because there is an overt subject beyond the relative clause. This applies also to longer constructions (20).

(20) mtoto wangu mzuri yule ninayempenda sana anasoma
```
mtoto
       "mtoto" N 1/2-SG { the } { child } @SUBJ
wangu
```

```
      "angu" PRON POSS 1/2-SG SG1 { my } @GCON
mzuri
      "zuri" ADJ A-INFL 1/2-SG { good } @<NADJ
yule
      "yule" PRON DEM :le 1/2-SG { that }
ninayempenda
      "penda" V 1/2-SG1-SP VFIN { *i } PR:na 1/2-SG-OBJ-REL { whom
} 1/2-SG3-OBJ OBJ NO-OBJ-GLOSS [penda] { love } SVO @FMAINVtr-OBJ>
sana
      "sana" AD-ADJ AR { much }
anasoma
      "soma" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [soma] { read }
SVO @FMAINVtr-OBJ>
```

3.2 Insertion

There are also cases where lexical items need to be inserted into the target text, although insertion is less prominent than deletion. A typical case is the article attached to the noun. Because Swahili does not have an article or any other way of marking definiteness, the management of the article is problematic.

In the morphological analyzer, each noun is given a definite article by default. Therefore, if no rule for manipulating the article applies, the noun is considered definite and preceded by the article 'the'. Whether the article should be definite or indefinite is very difficult to implement, because Swahili does not give any clue for that.

In the current implementation, the article is inserted by default into each noun, and its presence is then manipulated using re-writing rules. For example, the article is removed when the noun has a possessive or numerical attribute. Consider the examples in (21a) and (21b).

```
(20)
 (21a) mtoto wangu
mtoto
      "mtoto" N 1/2-SG { the } { child }
wangu
      "angu" PRON POSS 1/2-SG SG1 { my } @GCON

 (21b)
( PRON POSS 1/2-SG SG1 { my } ) ( N 1/2-SG { child } )
```

The article present in (21a) is removed in (21b). Also the word order has changed. Similar processes are needed also for numerical attributes (22a and 22b).

```
(22a) watoto watatu
watoto
      "mtoto" N 1/2-PL { the } { child }
watatu
      "tatu" NUM 1/2-PL NUM-INFL CARD { three }
```

(22b)
```
( NUM 1/2-PL NUM-INFL CARD { three } ) ( N 1/2-PL  { child } )
```

An example of a rather complicated relative structure, where insertion is involved, is in (23). The relative is composed of two units, 'ambaye' and 'naye', with a finite verb in between. The structure corresponds to the relative pronoun with a preposition in English (e.g. 'with whom', 'with which'). The disambiguated version is in (23a). Note that the presence or absence of the subject prefix gloss is already controlled in it. Also the role of 'ambaye' as an object is recognized. The pronoun 'naye', followed by a finite verb with an agreeing subject prefix, would be a subject, but here 'naye' is part of the relative structure and cannot be a subject.

In (23b), the lexical constituents in TL have been organized in the way they should be. This involves reordering, deletion and insertion.

In (23c), the surface form of the TL is presented. Note that the lexical glosses have been converted to appropriate surface forms.

(23)
```
 (23a) mtoto ambaye nilikaa naye amesafiri
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP @SUBJ
ambaye
      "ambaye" PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { whom }
nilikaa
      "kaa" V 1/2-SG1-SP VFIN { *i } PAST [kaa] { stay } SVO
@FMAINVtr+OBJ>
naye
      "naye" CC PRON PERS CC-SG 1/2-SG { with him/her }
amesafiri
      "safiri" V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me [safiri] {
travel } SV AR @FMAINVintr
```

(23b)
```
( N 1/2-SG { the } { child } CAP @SUBJ )
( PRON **OBJ-CLB OBJ-GEN-REL 1/2-SG { with } { whom } )
( V 1/2-SG1-SP VFIN { *i } PAST { stay } SVO )
( CC PRON CC-SG 1/2-SG  )
( V 1/2-SG3-SP VFIN NO-SP-GLOSS PERF:me { travel } SV )
```

(23c)
the child with whom I stayed has travelled.

In (24), there is an example of the corresponding relative structure with a plural non-animate relative referent.

(24)
```
 (24a) vitu ambavyo nilikaa navyo vimepotea
*vitu
      "kitu" N 7/8-PL { the } { thing } CAP @SUBJ
```

```
ambavyo
       "ambavyo" PRON **SUBJ-CLB OBJ-GEN-REL 7/8-PL { which }
nilikaa
       "kaa" V 1/2-SG1-SP VFIN { *i } PAST [kaa] { stay } SVO
@FMAINVtr-OBJ>
navyo
       "navyo" CC PRON CC-PL 7/8-PL { with them }
vimepotea
       "potea" V 7/8-PL-SP VFIN NO-SP-GLOSS PERF:me [potea] { be
lost } SVO @FMAINVtr-OBJ>
```

(24b)
```
( N 7/8-PL { the } { thing } CAP @SUBJ )
( PRON **SUBJ-CLB OBJ-GEN-REL 7/8-PL { with } { which } )
( V 1/2-SG1-SP VFIN { *i } PAST { stay } SVO )
( CC PRON CC-PL 7/8-PL  )
( V 7/8-PL-SP VFIN NO-SP-GLOSS PERF:me { be lost } SVO )
```

(24c)
the things with which I stayed have been lost

3.3 Substitution

A fairly common process in MT is substitution. Also here we talk about processes on the lexical level. An example of substitution is in (25).

(25) Watoto niliokuwa nao wamesafiri.
```
*watoto
       "mtoto" N 1/2-PL { the } { child } CAP
niliokuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST 1/2-PL-OBJ-REL { whom }
INFMARK [wa] { be } AUX-WA SV
nao
       "nao" CC PRON PERS CC-PL 1/2-PL { with them }
wamesafiri
       "safiri" V 1/2-PL3-SP VFIN { they } PERF:me [safiri] {
travel } SV AR
```

In (25), the words 'niliokuwa' and 'nao' form a multi-word unit expressing possession. Because the unit inflects, it must be treated so that all grammatically correct forms become possible. The substitution process needed is: be + with them > have. This is implemented in (26).

(26)
```
*watoto
       "mtoto" N 1/2-PL { the } { child } CAP
niliokuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST 1/2-PL-OBJ-REL { whom }
INFMARK [wa] IDOM-V> AUX-WA SV
```

```
nao
       "nao" <IDIOM { have }
wamesafiri
       "safiri" V 1/2-PL3-SP VFIN { they } PERF:me [safiri] {
travel } SV AR
```

In the next phase, the alternative readings in regard to the presence of the subject prefix gloss are introduced (27).

```
(27)
*watoto
       "mtoto" N 1/2-PL { the } { child } CAP @SUBJ+rel
niliokuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST 1/2-PL-OBJ-REL { whom } INFMARK
AUX-WA SV IDIOM-V> @FMAINVintr
       "wa" V 1/2-SG1-SP VFIN NO-SP-GLOSS PAST 1/2-PL-OBJ-REL { whom }
INFMARK AUX-WA SV IDIOM-V> @FMAINVintr
nao
       "nao" <IDIOM { have }
wamesafiri
       "safiri" V 1/2-PL3-SP VFIN { they } PERF:me [safiri] { travel } SV
AR @FMAINVintr
       "safiri" V 1/2-PL3-SP VFIN NO-SP-GLOSS PERF:me [safiri] { travel }
SV AR @FMAINVintr
```

Then the correct readings are chosen using CG rules (28).

```
(28)
*watoto
       "mtoto" N 1/2-PL { the } { child } CAP @SUBJ+rel
niliokuwa
       "wa" V 1/2-SG1-SP VFIN { *i } PAST 1/2-PL-OBJ-REL { whom }
INFMARK AUX-WA SV IDIOM-V> @FMAINVintr
nao
       "nao" <IDIOM { have }
wamesafiri
       "safiri" V 1/2-PL3-SP VFIN NO-SP-GLOSS PERF:me [safiri] {
travel } SV AR @FMAINVintr
```

Now when deletion, insertion and substitution operations have been performed on the lexical level, we still have to reorder the constituents to meet the requirements of the TL (29).

```
(29)
( N 1/2-PL { the } { child } CAP @SUBJ+rel )
( V 1/2-SG1-SP { whom } PAST 1/2-PL-OBJ-REL VFIN { *i }  INFMARK
AUX-WA SV IDIOM-V> <IDIOM { have } )
( V 1/2-PL3-SP VFIN NO-SP-GLOSS PERF:me { travel } SV )
```

When all operations on the lexical level have been performed, we can produce surface forms using the grammatical information inherited from the SL (30).

(30)
```
( N 1/2-PL { the } { children } CAP @SUBJ+rel )
( V 1/2-SG1-SP { whom } PAST 1/2-PL-OBJ-REL VFIN { *i }  INFMARK
AUX-WA SV IDIOM-V> <IDIOM { had } )
( V 1/2-PL3-SP VFIN NO-SP-GLOSS PERF:me { have } { travelled } SV
)
```

3.4 Combination of operations

An illuminating example of how various operations have to be applied is the translation of such relative constructions, where the relative refers to possession. English has a unique genitive form for the relative pronoun, (e.g. 'the mother whose child'), when the referent is animate. In the case of non-animate relative pronouns, a different construction is used (e.g. 'the tree the leaves of which...'). In Swahili, the construction is the same with animate and non-animate referents. The processing of such constructions is demonstrated below. Consider first the example with an animate referent (31).

(31) Mtoto ambaye mkono wake umevunjika analia.
```
*mtoto
      "mtoto" N 1/2-SG { the } { child } CAP @SUBJ+rel
ambaye
      "ambaye" PRON **POS-CLB POS-GEN-REL REL 1/2-SG { whose }
mkono
      "mkono" N 3/4-SG { the } { hand } @SUBJ
wake
      "ake" PRON POSS 3/4-SG SG3 { his/her/its } @GCON
umevunjika
      "vunjika" V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me [vunja] { get
broken } PREFR SV STAT @FMAINVintr
analia
      "lia" V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na [lia] { cry } SVO
@FMAINVtr-OBJ>
```

In (31), the relative pronoun 'ambaye' has a tag POS-GEN-REL meaning that it is a possessive general relative form. It already has the genitive gloss 'whose'. As a result of disambiguation, both of the finite verbs have the tag NO-SP-GLOSS, the first verb because of the overt subject (@SUBJ), and the second because of the overt subject (@SUBJ+rel) beyond the relative clause. It is still needed that the article gloss in 'mkono' is deleted, the gloss in 'ake' is deleted, and the order of constituents is modified. The result of these operations is demonstrated in (32).

(32)
```
( N 1/2-SG { the } { child } CAP @SUBJ+rel )
( PRON **POS-CLB POS-GEN-REL REL 1/2-SG { whose } )
( PRON POSS 3/4-SG SG3 )
( N 3/4-SG { hand } @SUBJ )
```

```
( V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me { get broken } PREFR SV
STAT )
( V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na { cry } SVO )
```

The surface forms of the TL are produced in (33).

(33)
```
( N 1/2-SG { the } { child } CAP @SUBJ+rel )
( PRON **POS-CLB POS-GEN-REL REL 1/2-SG { whose } )
:( PRON POSS 3/4-SG SG3 )
:( N 3/4-SG { hand } @SUBJ )
( V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me { has } { got broken }
PREFR SV STAT )
( V 1/2-SG3-SP VFIN NO-SP-GLOSS PR:na { cries } SVO )
```

The child whose hand has got broken cries

The processes in cases with a non-animate referent are more complicated. Consider the disambiguated example in (34).

(34) Mti ambao majani yake yamekauka umeanguka.
```
*mti
      "mti" N 3/4-SG { the } { tree } PLACE CAP @SUBJ+rel
ambao
      "ambao" PRON POS-GEN-REL REL 3/4-SG { which }
majani
      "jani" N 5/6-PL { the } { leaf } @SUBJ
yake
      "ake" PRON POSS 5/6-PL SG3 { his/her/its } @GCON
yamekauka
      "kauka" V 5/6-PL-SP VFIN NO-SP-GLOSS PERF:me [kauka] { dry
up } SV EXT: STAT :EXT @FMAINVintr
umeanguka
      "anguka" V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me [angua] { fall
down } PREFR SVO STAT @FMAINVtr-OBJ>
```

This case requires deletion, insertion, and fairly complicated reordering. The result is in (35).

(35)
```
( N 3/4-SG { the } { tree } PLACE CAP @SUBJ+rel )
( PRON POS-GEN-REL REL 3/4-SG )
( N 5/6-PL  { the } { leaf } @SUBJ )
( { of } { which } )
( PRON POSS 5/6-PL SG3 )
( V 5/6-PL-SP VFIN NO-SP-GLOSS PERF:me { dry up } SV STAT )
( V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me { fall down } PREFR SVO
STAT )
```

The version with correct surface forms in TL is in (36)

```
(36)
( N 3/4-SG { the } { tree } PLACE CAP @SUBJ+rel )
( PRON POS-GEN-REL REL 3/4-SG )
( N 5/6-PL  { the } { leaves } @SUBJ )
( { of } { which } )
:( PRON POSS 5/6-PL SG3 )
( V 5/6-PL-SP VFIN NO-SP-GLOSS PERF:me { have } { dried up } SV
STAT )
( V 3/4-SG-SP VFIN NO-SP-GLOSS PERF:me { has } { fallen down }
PREFR SVO STAT )
```

The tree the leaves of which have dried up has fallen down

4 Discussion and conclusion

In this report I have discussed various problems in translating relative constructions from Swahili into English. I have also discussed various alternatives for solving the problems and demonstrated phase by phase some solutions. It has become clear from the above that a problem can be solved in more than one way, some of which are more optimal than others. Therefore, it is worth considering the optimality of solutions, although in many cases the decision is difficult.

The choice is often between two alternatives.

(a) Should we make use of the extensive constraint alternatives of the CG parser - whereby we have to increase ambiguity in the morphological analyzer - or

(b) could we do the required modifications using re-writing rules that are applied after disambiguation. The implementation discussed above uses both methods.

For example, alternative glosses for words are written in the morphological lexicon, and the CG parser selects the appropriate gloss in each case. The insertion of glosses could also be done after the morphological analysis.

On the other hand, the tags for controlling the presence or absence of glosses for subject prefixes and object prefixes are inserted after initial disambiguation, and another set of CG rules is applied for controlling the correct readings. The reasons for using this solution are (a) the need of avoiding unnecessary complexity of the morphological analyzer, and (b) the use of the power of the CG grammar. Putting the ambiguity already in the morphological parser would be possible, but it would unnecessarily burden morphological analysis and disambiguation. The operations could also be done afterwards using re-write rules. However, this would be difficult because of limited possibilities of writing constraints.

Although rule-based MT is labour-intensive and requires the mastery of the grammar of the SL and TL, it makes use of knowledge on a deeper level, not on the surface only. Therefore, it should be considered the only reliable method for constructing language applications, such as MT.

In conclusion, the translation given by SALAMA to the examples discussed above is in Appendix 1.

Appendix 1

The examples discussed in this report were translated with SALAMA. Below is the translation of each example as given by the system.

1a. Mtu anayesoma.
The man who reads.

1b. Mtu asomaye.
The man who reads.

1c. Mtu ambaye anasoma.
The man who reads.

2a. Huyu ni mtu anayesoma.
This is the man who reads.

2b. Huyu ni mtu asomaye.
This is the man who reads.

2c. Huyu ni mtu ambaye anasoma.
This is the man who reads.

3a. Mtu anayesoma sana anafaulu katika maisha.
The man who reads much is successful in the life.

3b. Mtu asomaye sana anafaulu katika maisha.
The man who reads much is successful in the life.

3c. Mtu ambaye anasoma sana anafaulu katika maisha.
The man who reads much is successful in the life.

4a. Anayesoma sana anafaulu katika maisha.
Who reads much is successful in the life.

4b. Asomaye sana anafaulu katika maisha.
Who reads much is successful in the life.

4c. Ambaye anasoma sana anafaulu katika maisha.
Who reads much is successful in the life.

5. Mtoto anayempenda mama yake atafaulu.

The child who loves his/her/its mother will be successful.

6. Mtoto ninayempenda amesafiri.
The child whom I love has travelled.

7a. Mtoto niliye naye amesafiri.
The child whom I have has travelled.

7b. Mtoto niliyekuwa naye amesafiri.
The child whom I had has travelled.

7c. Mtoto ambaye nilikuwa naye amesafiri.
The child whom I had has travelled.

7d. Mtoto ninayekaa naye amesafiri.
The child with whom I stay has travelled.

7e. Mtoto ambaye ninakaa naye amesafiri.
The child with whom I stay has travelled.

8a. Mtoto nisiye naye amesafiri.
The child whom I did not have has travelled.

8b. Mtoto nisiyekuwa naye amesafiri.
The child whom I do not have has travelled.

8c. Mtoto ambaye sikuwa naye amesafiri.
The child whom I did not have has travelled.

8d. Mtoto nisiyekaa naye amesafiri.
The child with whom I do not stay has travelled.

8e. Mtoto ambaye sikai naye amesafiri.
The child with whom I do not stay has travelled.

9. Mtoto ambaye mkono wake umevunjika analia.
The child whose hand has got broken cries.

10. Mti ambao majani yake yamekauka umeanguka.
The tree the leaves of which have dried up has fallen down.

11. Mtu anayesoma.
The man who reads.

12. Anayesoma.
Who reads.

13. Mtoto anayempenda mama yake.
The child who loves his/her/its mother.

14. Mtoto anayempenda.
The child who loves him/her.

15. Anayempenda.
Who loves him/her.

16. Mtoto anayesoma kitabu.
The child who reads the book.

17. Mtoto anayempenda mama anasoma.
The child who loves the mother reads.

18. Mtoto anayempenda mama anasoma.
The child who loves the mother reads.

19. Mtoto ninayempenda anasoma.
The child whom I love reads.

20. Mtoto wangu mzuri yule ninayempenda sana anasoma.
That my good child whom I love much reads.

21. Mtoto wangu.
My child.

22. Watoto watatu.
Three children.

23. Mtoto ambaye nilikaa naye amesafiri.
The child with whom I stayed has travelled.

24. Vitu ambavyo nilikaa navyo vimepotea.
The things with which I stayed have been lost.

25. Watoto niliokuwa nao wamesafiri.
The children whom I had have travelled.

31. Mtoto ambaye mkono wake umevunjika analia.
The child whose hand has got broken cries.

34. Mti ambao majani yake yamekauka umeanguka.
The tree the leaves of which have dried up has fallen down.