

Constructing pronoun forms in English to Swahili machine translation

Arvi Hurskainen
Department of World Cultures, Box 59
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi

Abstract

It is hard to find a linguistic feature, where English and Swahili are identical. While English verb has only five possible forms, Swahili verb has theoretically millions of valid forms. Not all of them can be found in practice for each verb lemma. However, it is not difficult to find over two thousand forms for a single verb. The complexity derives from the noun class system and from the fact that Swahili encodes morphological and also syntactic features as part of the verb. Most of these features are prefixes, but also verb derivation by means of suffixes is common. In this report we investigate one specific problem deriving from differences in language structure. We try to find out how Swahili pronoun forms can be produced in cases, where the referent of the pronoun is difficult to identify.

Key Words: *pronoun forms, machine translation*

1 Introduction

The morphologically and syntactically analysed text contains all the relevant linguistic information that can be derived from the text. The result is a structured and classified representation of the surface text. When we consider translating the text into another language, we only have the information of the source text. English and Swahili are vastly different languages, and direct lexical transfer is not possible. In order to achieve correct translation, we must take into account the requirements of the target language.

These specific requirements include the noun class system, which is missing in the source language. For example, English demonstrative pronouns have only singular and plural forms. Swahili has also singular and plural forms, but each of them has a separate form for each noun class. Therefore, while a demonstrative pronoun in English has two forms, singular and plural, in Swahili it has 15 forms. The big number of forms is not a problem in itself, but problems emerge when we have to assign the correct noun class for the pronoun, when the source language does not give any clue. The information must be found somewhere else in the sentence.

Where can the needed information be found, so that the correct pronoun form can be constructed? The information is normally available somewhere in the context. When lexical translations for each word are transferred to the analysed text, nouns have their noun class definitions. This already is a rich source for handling noun class specifications. There is also some semantic classification that helps in assigning correct noun classes.

When the analysed text of English is enriched with Swahili glosses, it is possible to add tags to words, so that the correct form of each word can finally be produced. Here we concentrate on pronouns.

2 Adding tags to demonstrative pronouns

The translation of pronouns is often problematic, because we must assign the noun class to the pronoun, although English does not provide any clue to what it should be. In case the pronoun is a modifier of the noun, the problem is easy to solve, because the noun head defines it. Examples are in (1).

(1)

This man is tall.

Those trees are tall.

My trees are tall.

Translation:

Mwanadamu huyu ni mkubwa.

Miti ile ni mikubwa.

Miti yangu ni mikubwa.

The key for defining the tag is on the right

The task becomes more problematic, when the demonstrative pronoun alone is a subject, such as in (2).

(2)

a. *This is a man.*

b. *These are men.*

c. *This is a tree.*

d. *These are trees.*

e. *This is a hoe.*

f. *These are hoes.*

g. *This is a chair.*

h. *These are chairs.*

i. *This is a lamp.*

j. *These are lamps.*

k. *This is freedom.*

l. *This is a place.*

These examples represent various noun classes. In English the pronoun is either singular (this) or plural (these), but in Swahili we must add also the correct noun class form. The task is more complex, because the needed information can be far from the pronoun. For the sake of clarity, the examples in (2) are as short as possible. After the morphological analysis, morphological disambiguation, syntactic mapping, lexical transfer of Swahili, and semantic disambiguation, the result is as in (3).

(3)

a.

```
"<This>"
    "this" { h } %SUBJ CAPINIT PRON DEM SG
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<a>"
    "a" %DN> DET SG
"<man>"
    "man" { 1SG 2PL anadamu } AN %PCOMPL-S N SG NOM INDEF
"<.>"
    "."
```

b.

```
"<These>"
    "this" { h } %SUBJ CAPINIT PRON DEM PL
"<are>"
    "be" { ni } MONOSLB %+FMAINV V PRES PL
"<men>"
    "man" { 1SG 2PL anadamu } AN %PCOMPL-S N PL NOM INDEF
"<.>"
    "."
```

c.

```
"<This>"
    "this" { h } %SUBJ CAPINIT PRON DEM SG
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<a>"
    "a" %DN> DET SG
"<tree>"
    "tree" { 3SG 4PL ti } %PCOMPL-S N SG NOM INDEF
"<.>"
    "."
```

d.

```
"<These>"
    "this" { h } %SUBJ CAPINIT PRON DEM PL
"<are>"
    "be" { ni } MONOSLB %+FMAINV V PRES PL
"<trees>"
    "tree" { 3SG 4PL ti } %PCOMPL-S N PL NOM INDEF
"<.>"
    "."
```

e.

```
"<This>"
    "this" { h } %SUBJ CAPINIT PRON DEM SG
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<a>"
    "a" %DN> DET SG
"<hoe>"
    "hoe" { 5SG 6PL jembe } %PCOMPL-S N SG NOM INDEF
```

```
"<.>"  
    "."  
f.  
"<These>"  
    "this" { h } %SUBJ CAPINIT PRON DEM PL  
"<are>"  
    "be" { ni } MONOSLB %+FMAINV V PRES PL  
"<hoes>"  
    "hoe" { 5SG 6PL jembe } %PCOMPL-S N PL NOM INDEF  
"<.>"  
    "."  
g.  
"<This>"  
    "this" { h } %SUBJ CAPINIT PRON DEM SG  
"<is>"  
    "be" { ni } MONOSLB %+FMAINV V PRES SG3  
"<a>"  
    "a" %DN> DET SG  
"<chair>"  
    "chair" { 7SG 8PL ti } %PCOMPL-S N SG NOM INDEF  
"<.>"  
    "."  
h.  
"<These>"  
    "this" { h } %SUBJ CAPINIT PRON DEM PL  
"<are>"  
    "be" { ni } MONOSLB %+FMAINV V PRES PL  
"<chairs>"  
    "chair" { 7SG 8PL ti } %PCOMPL-S N PL NOM INDEF  
"<.>"  
    "."  
i.  
"<This>"  
    "this" { h } %SUBJ CAPINIT PRON DEM SG  
"<is>"  
    "be" { ni } MONOSLB %+FMAINV V PRES SG3  
"<a>"  
    "a" %DN> DET SG  
"<lamp>"  
    "lamp" { 9SG 10PL taa } %PCOMPL-S N SG NOM INDEF  
"<.>"  
    "."  
j.  
"<These>"  
    "this" { h } %SUBJ CAPINIT PRON DEM PL  
"<are>"  
    "be" { ni } MONOSLB %+FMAINV V PRES PL  
"<lamps>"  
    "lamp" { 9SG 10PL taa } %PCOMPL-S N PL NOM INDEF  
"<.>"  
    "."
```

k.
"<This>"
 "this" { h } %SUBJ CAPINIT PRON DEM SG
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<freedom>"
 "freedom" { 11SG huru } %PCOMPL-S N SG NOM INDEF
"<.>"
 "."

l.
"<This>"
 "this" { h } %SUBJ CAPINIT PRON DEM SG
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<a>"
 "a" %DN> DET SG
"<place>"
 "place" { 16SG mahali } %PCOMPL-S N SG NOM INDEF
"<.>"
 "."

We see that in all sentences the gloss for the demonstrative pronoun is *h*. This is the only stable element in the demonstrative pronoun. Now we should provide each pronoun with a tag, which would help in constructing the correct form for each pronoun. Tags are added for demonstrative pronouns in (4).

(4)

a.
"<This>"
 "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-1
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<man>"
 "man" { 1SG anadamu } AN %PCOMPL-S N SG NOM INDEF
"<.>"
 "."

b.
"<These>"
 "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-2
"<are>"
 "be" { ni } MONOSLB %+FMAINV V PRES PL
"<men>"
 "man" { 2PL anadamu } AN %PCOMPL-S N PL NOM INDEF
"<.>"
 "."

c.
"<This>"
 "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-3
"<is>"

```
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<tree>"
  "tree" { 3SG ti } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
d.
"<These>"
  "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-4
"<are>"
  "be" { ni } MONOSLB %+FMAINV V PRES PL
"<trees>"
  "tree" { 4PL ti } %PCOMPL-S N PL NOM INDEF
"<.>"
  "."
e.
"<This>"
  "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-5
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<hoe>"
  "hoe" { 5SG jembe } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
"<<s>>" "<s>"
f.
"<These>"
  "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-6
"<are>"
  "be" { ni } MONOSLB %+FMAINV V PRES PL
"<hoes>"
  "hoe" { 6PL jembe } %PCOMPL-S N PL NOM INDEF
"<.>"
  "."
g.
"<This>"
  "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-7
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<chair>"
  "chair" { 7SG ti } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
h.
"<These>"
  "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-8
"<are>"
  "be" { ni } MONOSLB %+FMAINV V PRES PL
"<chairs>"
  "chair" { 8PL ti } %PCOMPL-S N PL NOM INDEF
"<.>"
  "."
```

```
"<<s>>" "<s>"
i.
"<This>"
  "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-9
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<lamp>"
  "lamp" { 9SG taa } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
j.
"<These>"
  "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-10
"<are>"
  "be" { ni } MONOSLB %+FMAINV V PRES PL
"<lamps>"
  "lamp" { 10PL taa } %PCOMPL-S N PL NOM INDEF
"<.>"
  "."
k.
"<This>"
  "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-11
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<freedom>"
  "freedom" { 11SG huru } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
l.
"<This>"
  "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-16
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<place>"
  "place" { 16SG mahali } %PCOMPL-S N SG NOM INDEF
"<.>"
  "."
```

In all these examples, the information on assigning the inflection tag for demonstrative pronouns is found from the post complement noun of the sentence.

In the next phase the tag is attached to the Swahili gloss (5).

(5)

```
a.
"<This>"
  "this" { h+DEM-1 } %SUBJ CAPINIT PRON DEM SG
"<is>"
  "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<man>"
  "man" { 1SG anadamu } AN %PCOMPL-S N SG NOM INDEF
"<.>"
```

"."

b.
"<These>"
"this" { h+DEM-2 } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<men>"
"man" { 2PL anadamu } AN %PCOMPL-S N PL NOM INDEF
"<.>"
"."

c.
"<This>"
"this" { h+DEM-3 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<tree>"
"tree" { 3SG ti } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

d.
"<These>"
"this" { h+DEM-4 } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<trees>"
"tree" { 4PL ti } %PCOMPL-S N PL NOM INDEF
"<.>"
"."

e.
"<This>"
"this" { h+DEM-5 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<hoe>"
"hoe" { 5SG jembe } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

f.
"<These>"
"this" { h+DEM-6 } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<hoes>"
"hoe" { 6PL jembe } %PCOMPL-S N PL NOM INDEF
"<.>"
"."

g.
"<This>"
"this" { h+DEM-7 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3

"<chair>"
"chair" { 7SG ti } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

h.
"<These>"
"this" { h+DEM-8 } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<chairs>"
"chair" { 8PL ti } %PCOMPL-S N PL NOM INDEF
"<.>"
"."

i.
"<This>"
"this" { h+DEM-9 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<lamp>"
"lamp" { 9SG taa } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

j.
"<These>"
"this" { h+DEM-10 } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<lamps>"
"lamp" { 10PL taa } %PCOMPL-S N PL NOM INDEF
"<.>"
"."

k.
"<This>"
"this" { h+DEM-11 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<freedom>"
"freedom" { 11SG huru } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

l.
"<This>"
"this" { h+DEM-16 } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<place>"
"place" { 16SG mahali } %PCOMPL-S N SG NOM INDEF
"<.>"
"."

Then we convert the tag to surface form. Also other morphological tags are converted into surface form (6).

(6)
a.
"<This>"
 "this" { h+uyu } %SUBJ CAPINIT PRON DEM SG
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<man>"
 "man" { mw+anadamu } AN %PCOMPL-S N SG NOM INDEF
"<.>"
 "."
b.
"<These>"
 "this" { h+awa } %SUBJ CAPINIT PRON DEM PL
"<are>"
 "be" { ni } MONOSLB %+FMAINV V PRES PL
"<men>"
 "man" { w+anadamu } AN %PCOMPL-S N PL NOM INDEF
"<.>"
 "."
c.
"<This>"
 "this" { h+uu } %SUBJ CAPINIT PRON DEM SG
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<tree>"
 "tree" { m+ti } %PCOMPL-S N SG NOM INDEF
"<.>"
 "."
d.
"<These>"
 "this" { h+ii } %SUBJ CAPINIT PRON DEM PL
"<are>"
 "be" { ni } MONOSLB %+FMAINV V PRES PL
"<trees>"
 "tree" { mi+ti } %PCOMPL-S N PL NOM INDEF
"<.>"
 "."
e.
"<This>"
 "this" { h+ili } %SUBJ CAPINIT PRON DEM SG
"<is>"
 "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<hoe>"
 "hoe" { jembe } %PCOMPL-S N SG NOM INDEF
"<.>"
 "."
f.
"<These>"

```
"this" { h+aya } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<hoes>"
"hoe" { ma+jembe } %PCOMPL-S N PL NOM INDEF
"<.>"
"."
g.
"<This>"
"this" { h+iki } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<chair>"
"chair" { ki+ti } %PCOMPL-S N SG NOM INDEF
"<.>"
"."
h.
"<These>"
"this" { h+ivi } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<chairs>"
"chair" { vi+ti } %PCOMPL-S N PL NOM INDEF
"<.>"
"."
i.
"<This>"
"this" { h+ii } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<lamp>"
"lamp" { taa } %PCOMPL-S N SG NOM INDEF
"<.>"
"."
j.
"<These>"
"this" { h+izi } %SUBJ CAPINIT PRON DEM PL
"<are>"
"be" { ni } MONOSLB %+FMAINV V PRES PL
"<lamps>"
"lamp" { taa } %PCOMPL-S N PL NOM INDEF
"<.>"
"."
k.
"<This>"
"this" { h+uu } %SUBJ CAPINIT PRON DEM SG
"<is>"
"be" { ni } MONOSLB %+FMAINV V PRES SG3
"<freedom>"
"freedom" { u+huru } %PCOMPL-S N SG NOM INDEF
"<.>"
```

```
    "."  
1.  
"<This>"  
    "this" { h+apa } %SUBJ CAPINIT PRON DEM SG  
"<is>"  
    "be" { ni } MONOSLB %+FMAINV V PRES SG3  
"<place>"  
    "place" { mahali } %PCOMPL-S N SG NOM INDEF  
"<.>"  
    "."
```

The final translation is in (7).

(7)

- a. *Huyu ni mwanadamu.*
- b. *Hawa ni wanadamu.*
- c. *Huu ni mti.*
- d. *Hii ni miti.*
- e. *Hili ni jembe.*
- f. *Haya ni majembe.*
- g. *Hiki ni kiti.*
- h. *Hivi ni viti.*
- i. *Hii ni taa.*
- j. *Hizi ni taa.*
- k. *Huu ni uhuru.*
- l. *Hapa ni mahali.*

3 The key for defining the tag is on the left

In the sentences discussed above, the needed information is found on the first noun on the right, in the same clause. The problem becomes much more difficult, if the demonstrative pronoun refers to a noun somewhere on the left, in another sentence.

One approach is to scan to the left for finding a noun, which the pronoun would refer to. Most likely the correct pronoun would be in the object position. If object is not found in the previous sentence, the next likely noun would be the subject. However, the scanning should be stopped on the paragraph boundary. With these considerations in mind, we can test with some examples.

The Constraint Grammar rule is of the type as in (8).

(8)

```
MAP (DEM-3) TARGET DEM (*-1 (3SG) + N BARRIER PB);
```

This reads: *Map the demonstrative pronoun (DEM) with the tag DEM-3, if somewhere on the left there is the tag 3SG plus N (referring to noun). Do not scan beyond paragraph boundary (PB).*

This rule would work, if the referent of the pronoun is in the previous sentence and there are no other nouns with different noun class.

Note that there is a rule for each noun class, and each of them would apply the rule if conditions are fulfilled. Therefore, one must be very careful in designing the rules.

One solution would be to construct subsequent rules as in (9).

(9)

(a) MAP (DEM-3) TARGET DEM + SG (*-1 (3SG) + OBJ BARRIER OBJ OR PB);

(b) MAP (DEM-3) TARGET DEM + SG (*-1 (3SG) + SUBJ BARRIER SUBJ OR PB);

(c) MAP (DEM-3) TARGET DEM + SG (*-1 (3SG) + N BARRIER PB);

Rule (a) would look for the tag 3SG of the noun in object position on the left side. If the first object has this tag, the rule applies. Paragraph boundary is the last point of scanning.

If object with the tag 3SG is not found, the rule (b) is tried. If the first subject on the left has the tag 3SG, the rule is applied. Also here the scanning is stopped on paragraph boundary.

The last alternative is rule (c). It allows scanning to the left until paragraph boundary for finding the tag 3SG. The rule is very unreliable and unlikely to apply in real text.

In (10) is the set of rules for controlling the pronoun tag in all noun class cases. The rules combine the features of the rules in (9) above.

(10)

MAP (DEM-1) TARGET DEM + SG (*-1 (1SG) + N OR (2PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-2) TARGET DEM + PL (*-1 (2PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-3) TARGET DEM + SG (*-1 (3SG) + N OR (4PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-4) TARGET DEM + PL (*-1 (4PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-5) TARGET DEM + SG (*-1 (5SG) + N OR (6PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-6) TARGET DEM + PL (*-1 (6PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-7) TARGET DEM + SG (*-1 (7SG) + N OR (7PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-8) TARGET DEM + PL (*-1 (8PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-9) TARGET DEM + SG (*-1 (9SG) + N OR (10PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-10) TARGET DEM + PL (*-1 (10PL) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-11) TARGET DEM + SG (*-1 (11SG) + N BARRIER OBJ OR SUBJ OR PB);

MAP (DEM-16) TARGET DEM + SG (*-1 (16SG) + N BARRIER OBJ OR SUBJ OR PB);

Below are some examples for testing how the rules work.

```
(11)
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<has>"
    "have" { -na } %+FMAINV V PRES SG3 OBJ-10 SP-1 SP-1
"<many>"
    "many" { ingi } %QN> DET ABS PL I-4
"<trees>"
    "tree" { 4PL ti } %OBJ N PL
"<.>"
    "."
"<This>"
    "this" { h } %SUBJ CAPINIT PRON DEM SG DEM-3
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<very>"
    "very" { sana } %AD-A> ADV
"<tall>"
    "tall" { -kubwa } A-INFL %PCOMPL-S A ABS A-4
"<.>"
    "."
```

The pronoun tag DEM-3 was added, because the noun in object position in the previous sentence has the tag 4PL. Note that although the tag in the preceding sentence is a plural tag, the pronoun gets a singular tag. Both tags belong to the same semantic group. We can test the rules, when the pronoun is in plural (12).

```
(12)
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<has>"
    "have" { -na } %+FMAINV V PRES SG3 OBJ-10 SP-1 SP-1
"<many>"
    "many" { ingi } %QN> DET ABS PL I-4
"<trees>"
    "tree" { 4PL ti } %OBJ N PL
"<.>"
    "."
"<These>"
    "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-4
"<are>"
    "be" { ni } MONOSLB %+FMAINV V PRES PL
"<very>"
    "very" { sana } %AD-A> ADV
"<tall>"
    "tall" { -kubwa } A-INFL %PCOMPL-S A ABS A-9
"<.>"
    "."
```

Now the pronoun has been given a plural tag of the same semantic group.

Let us see what happens when we have more than one preding sentence with a noun in object position (13).

(13)

```
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<bought>"
    "buy" { nunuA } HUM-V SVO %+FMAINV V PAST TAM-li OBJ-10 SP-1
SP-1
"<lamps>"
    "lamp" { 10PL taa } %OBJ N PL INDEF
"<.>"
    "."
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<bought>"
    "buy" { nunuA } HUM-V SVO %+FMAINV V PAST TAM-li OBJ-10 SP-1
SP-1
"<hoes>"
    "hoe" { 6PL jembe } %OBJ N PL INDEF
"<.>"
    "."
"<These>"
    "this" { h } %SUBJ CAPINIT PRON DEM PL DEM-6
"<were>"
    "be" { INFMARK+wA } MONOSLB %+FMAINV V PAST PL COND-CAND
TAM-li
"<expensive>"
    "expensive" { ghali } UNINFL %PCOMPL-S A ABS INDEF
"<.>"
    "."
```

The rule has used the first object on the left for giving the tag DEM-6 to the pronoun. It is not clear whether the pronoun refers to the nearest object or to both objects, but normally it is understood in such structures that it refers to the nearest one.

If we want that the pronoun refers to more than one object, we normally make it clear in some way. One method is to put the objects into the same sentence as in (14).

(14)

```
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<bought>"
    "buy" { nunuA } HUM-V SVO %+FMAINV V PAST TAM-li OBJ-10 SP-1
SP-1
"<hoes>"
    "hoe" { 6PL jembe } %OBJ N PL INDEF
"<and>"
    "and" { na } %CC CC
```

```
"<chairs>"
    "chair" { 8PL ti } %OBJ N PL
"<.>"
    "."
"<Those>"
    "that" { le } %SUBJ CAPINIT PRON DEM PL DEM-8
"<were>"
    "be" { INFMARK+wA } MONOSLB %+FMAINV V PAST PL COND-CAND
TAM-li SP-10
"<expensive>"
    "expensive" { ghali } UNINFL %PCOMPL-S A ABS INDEF
"<.>"
    "."
```

The first sentence in (14) sentence has two objects, and the rule has taken the last object as reference in assigning the tag DEM-8 for the pronoun. This is one method to refer to multiple objects. Another method, and even a more preferable one, is to use the class 10 tag. Class group 9/10 is considered a class of everything else which do not naturally fit into any other class. We must add a new rule for these cases, and this rule should precede the other rules for pronouns. The rule is in (15).

```
(15)
MAP (DEM-10) TARGET DEM + PL (*-1 OBJ + N BARRIER PB LINK *-1 OBJ
BARRIER SNTB);
```

The rule reads: *Map the tag DEM-10 to the plural demonstrative pronoun, if on the left in the same paragraph there is a noun object and in the same sentence another object.*

An example is in (16).

```
(16)
"<He>"
    "he" { NOGLOSS } %SUBJ ACR CAPINIT PRON PERS NOM SG3
"<bought>"
    "buy" { nunuA } HUM-V SVO %+FMAINV V PAST TAM-li SP-1 SP-1
"<hoes>"
    "hoe" { 6PL jembe } %OBJ N PL INDEF
"<and>"
    "and" { na } %CC CC
"<chairs>"
    "chair" { 8PL ti } %OBJ N PL
"<.>"
    "."
"<Those>"
    "that" { le } %SUBJ CAPINIT PRON DEM PL DEM-10
"<were>"
    "be" { INFMARK+wA } MONOSLB %+FMAINV V PAST PL COND-CAND
TAM-li SP-10
"<expensive>"
```



```
"expensive" { ghali } UNINFL %PCOMPL-S A ABS INDEF  
"<.>"  
  "."
```

The final translation of the above examples is in (17).

(17)
[11]
Ana miti mingi.
Huu ni mkubwa sana.

[12]
Ana miti mingi.
Hii ni mikubwa sana.

[13]
Alinunua taa.
Alinunua majembe.
Haya yalikuwa ghali.

[14]
Alinunua majembe na viti.
Zile zilikuwa ghali.

If it is not clear, what is precisely the referent of the demonstrative noun, the convention is that class 9 is used in singular and class 10 in plural.

In the examples above, the referent of the demonstrative pronoun was an object. In (18) we have an example, where the referent is not an object or subject.

```
(18)  
"<There>"  
  "there" MW>  
"<were>"  
  "be" V <MW { kulikuwa na }  
"<many>"  
  "many" { ingi } %QN> DET ABS PL I-6  
"<fields>"  
  "field" { 6PL shamba } %PCOMPL-S N PL NOM  
"<.>"  
  "."  
"<Those>"  
  "that" { le } %SUBJ CAPINIT PRON DEM PL DEM-6  
"<were>"  
  "be" { INFMARK+wA } MONOSLB %+FMAINV V PAST PL COND-CAND  
TAM-li  
"<large>"  
  "large" { -kubwa } A-INFL %PCOMPL-S A ABS INDEF  
"<.>"
```

" . "

The rule, copied from (10) above:

MAP (DEM-6) TARGET DEM + PL (*-1 (6PL) + N BARRIER OBJ OR SUBJ OR PB) ;

The rule applies, because it does not require that the referent must be object or subject. If any noun with the appropriate noun specification is closer than the object referent or subject referent, the rule applies using its specification. In case an object or subject would be closer to the pronoun, its specification would be taken for the pronoun tag, as the rule defines.

The translation of (18) is in (19).

(19)

Kulikuwa na mashamba mengi.

Yale yalikuwa makubwa.

4 Conclusion

The report discusses the problem of assigning the correct form for a demonstrative pronoun in English to Swahili machine translation. The discussion is restricted to such cases, where the pronoun is not a modifier of a noun. The key for defining the correct form may be somewhere on the right in the same sentence. It may also be on the left in another sentence. Using carefully designed mapping rules it is possible to produce clean translation in all cases.