

<https://helda.helsinki.fi>

Translation of complex word-forms between three structurally different languages

Hurskainen, Arvi

University of Helsinki, Institute for Asian and African Studies
2019

Hurskainen , A 2019 ' Translation of complex word-forms between three structurally different languages ' Technical reports on language technology , no. 48 , University of Helsinki, Institute for Asian and African Studies , Helsinki . <

<http://www.njas.helsinki.fi/salama/translation-of-complex-word-forms-between-three-structurally-different-languages.p>

>

<http://hdl.handle.net/10138/310888>

cc_by_nc
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Translation of complex word-forms between three structurally different languages

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

Abstract

Translation between structurally very different languages is challenging, because there are several factors, which must be taken into consideration. In this report we test translation of an example sentence, where a single verb of Swahili must be translated with nine words. Again, when translating the same from English to Swahili, nine words must be compressed into one word. The languages considered are Swahili, English and Finnish.

We also test how Google Translate with neural approaches succeeds in translation.

Key Words: *morphology, machine translation.*

1 Introduction

Translation between English and Swahili might sometimes be a nightmare, because the ways of constructing words and sentences are very different. However, when using the rule-based approach to translation, even such problems can be solved. We demonstrate the translation process to both directions, using a single sentence *What I will make you think, each of you, is this*. The sentence as such may not appear in any text, but it is grammatical and can be used in demonstrating complex structures. We also test the translation from Swahili to Finnish, using the same sentence. Finally, we compare the translation results between SALAMA and Google Translate.

2 English to Swahili

The English text is first analysed, disambiguated and provided with syntactic tags (1).

(1)

1	What	what	@OBJ %NH PRON WH
2	I	i	subj:>3 @SUBJ %NH PRON PERS NOM SG1
3	will	will	v-ch:>4 @+FAUXV %AUX V AUXMOD
4	make	make	subj:>12 @-FMAINV %VA V INF
5	you	you	subj:>6 @OBJ %NH PRON PERS ACC
6	think	think	obj:>4 @-FMAINV %VA V INF
7	,	,	
8	each	each	obj:>6 @OBJ %NH PRON SG
9	of	of	mod:>8 @<NOM-OF %N< PREP

```
10  you  you  pcomp:>9    @<P %NH PRON PERS ACC
11  ,    ,
12  is   be   main:>0    @+FMAINV %VA V PRES SG3
13  this this comp:>12  @PCOMPL-S %NH PRON DEM SG
14  .    .
```

It is then modified suitable to be processed by FDG rules (2).

```
(2)
"<What>"
    "what" %OBJ CAPINIT PRON WH
"<I>"
    "i" %SUBJ PRON PERS NOM SG1
"<will>"
    "will" %+FAUXV V AUXMOD
"<make>"
    "make" %-FMAINV V INF
"<you>"
    "you" %OBJ PRON PERS ACC
"<think>"
    "think" %-FMAINV V INF
"<,>"
    " ,"
"<each>"
    "each" %OBJ PRON SG
"<of>"
    "of" %<NOM-OF PREP
"<you>"
    "you" %<P PRON PERS ACC
"<,>"
    " ,"
"<is>"
    "be" %+FMAINV V PRES SG3
"<this>"
    "this" %PCOMPL-S PRON DEM SG
"<.>"
    " ."
```

Swahili glosses are added to the readings (3).

```
(3)
"<What>"
    "what { nini , gani , amba , NOGLOSS }" %OBJ CAPINIT PRON WH
"<I>"
    "i { mimi , -angu , NOGLOSS }" %SUBJ PRON PERS NOM SG1
"<will>"
    "will { FUT , pendA }" %+FAUXV V AUXMOD
"<make>"
    "make { fanyA , NOGLOSS } SVO" %-FMAINV V INF
```

```
"<you>"
    "you { wewe , -ako , NOGLOSS }" %OBJ PRON PERS ACC
"<think>"
    "think { fikiriri , dhani , wazA , oneleA , angaliA , fikiriri ,
hesabu , hisabu , taamali } HUM-V SVO" %-FMAINV V INF
"<,>"
    " ,"
"<each>"
    "each { kila mmoja }" %OBJ PRON SG
"<of>"
    "of { -a }" %<NOM-OF PREP
"<you>"
    "you { wewe , -ako , NOGLOSS }" %<P PRON PERS ACC
"<,>"
    " ,"
"<is>"
    "be { INFMARK+wA , wA , ni , si , AUX , LOC } MONOSLB"
%+FMAINV V PRES SG3
"<this>"
    "this { h }" %PCOMPL-S PRON DEM SG
"<.>"
    " ."
```

This is reformatted, so that semantic disambiguation rules can be applied (4).

(4)

```
"<What>"
    "what" { nini } %OBJ CAPINIT PRON WH
    "what" { gani } %OBJ CAPINIT PRON WH
    "what" { amba } %OBJ CAPINIT PRON WH
    "what" { NOGLOSS } %OBJ CAPINIT PRON WH
"<I>"
    "i" { mimi } %SUBJ PRON PERS NOM SG1
    "i" { -angu } %SUBJ PRON PERS NOM SG1
    "i" { NOGLOSS } %SUBJ PRON PERS NOM SG1
"<will>"
    "will" { FUT } %+FAUXV V AUXMOD
    "will" { pendA } %+FAUXV V AUXMOD
"<make>"
    "make" { fanyA } SVO %-FMAINV V INF
    "make" { NOGLOSS } SVO %-FMAINV V INF
"<you>"
    "you" { wewe } %OBJ PRON PERS ACC
    "you" { -ako } %OBJ PRON PERS ACC
    "you" { NOGLOSS } %OBJ PRON PERS ACC
"<think>"
    "think" { fikiriri } HUM-V SVO %-FMAINV V INF
    "think" { dhani } HUM-V SVO %-FMAINV V INF
    "think" { wazA } HUM-V SVO %-FMAINV V INF
    "think" { oneleA } HUM-V SVO %-FMAINV V INF
    "think" { angaliA } HUM-V SVO %-FMAINV V INF
```

```
"<,>"  
    "," { , }  
"<each>"  
    "each" { kila mmoja } %OBJ PRON SG  
"<of>"  
    "of" { -a } %<NOM-OF PREP  
"<you>"  
    "you" { wewe } %<P PRON PERS ACC  
    "you" { -ako } %<P PRON PERS ACC  
    "you" { NOGLOSS } %<P PRON PERS ACC  
"<,>"  
    "," { , }  
"<is>"  
    "be" { INFMARK+wA } MONOSLB %+FMAINV V PRES SG3  
    "be" { wA } MONOSLB %+FMAINV V PRES SG3  
    "be" { ni } MONOSLB %+FMAINV V PRES SG3  
    "be" { si } MONOSLB %+FMAINV V PRES SG3  
    "be" { AUX } MONOSLB %+FMAINV V PRES SG3  
"<this>"  
    "this" { h } %PCOMPL-S PRON DEM SG  
"<.>"  
    "." { . }
```

When semantic disambiguation rules have been applied, we get the result as in (5).

```
(5)  
"<What>"  
    "what" { NOGLOSS } %OBJ CAPINIT PRON WH  
"<I>"  
    "i" { NOGLOSS } %SUBJ PRON PERS NOM SG1  
"<will>"  
    "will" { FUT } %+FAUXV V AUXMOD  
    "will" { penda } %+FAUXV V AUXMOD  
"<make>"  
    "make" { NOGLOSS } SVO %-FMAINV V INF  
"<you>"  
    "you" { NOGLOSS } %OBJ PRON PERS ACC  
"<think>"  
    "think" { fikiri } HUM-V SVO %-FMAINV V INF  
    "think" { dhani } HUM-V SVO %-FMAINV V INF  
    "think" { wazA } HUM-V SVO %-FMAINV V INF  
    "think" { oneleA } HUM-V SVO %-FMAINV V INF  
    "think" { angaliA } HUM-V SVO %-FMAINV V INF  
"<,>"  
    "," { , }  
"<each>"  
    "each" { kila mmoja } %OBJ PRON SG  
"<of>"  
    "of" { -a } %<NOM-OF PREP  
"<you>"  
    "you" { wewe } %<P PRON PERS ACC
```

```

    "you" { -ako } %<P PRON PERS ACC
    "you" { NOGLOSS } %<P PRON PERS ACC
"<,>"
    "," { , }
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<this>"
    "this" { h } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

We see above that five first words have a zero gloss. This means that those words are not translated with corresponding Swahili words. Their meaning is expressed with affixes in the verb. Also, the sequence *each of you* will not be translated. Its meaning will be expressed with an affix in the verb. We will see later how this is done. The verb *think* still has five readings. It needs no disambiguation, because the first one, *fikiri*, as default choice will be taken.

When we isolate multiword expressions, we see that there is one MWE in the sentence (6).

```

(6)
"<What>"
    "what" { NOGLOSS } %OBJ ACR CAPINIT PRON WH
"<I>"
    "i" { NOGLOSS } %SUBJ ACR PRON PERS NOM SG1
"<will>"
    "will" { FUT } %+FAUXV ACR V AUXMOD
"<make>"
    "make" { NOGLOSS } SVO %-FMAINV V INF
"<you>"
    "you" { NOGLOSS } %OBJ ACR PRON PERS ACC
"<think>"
    "think" { fikiri } HUM-V SVO %-FMAINV V INF
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { kila mmoja wenu } V MW
    "each_of_you" { IMP-SG2 } V MW
"<,>"
    "," { , }
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<this>"
    "this" { h } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

The MWE is disambiguated, and we see in (7) below, that the tag IMP-SG2 was selected instead of an overt gloss.

```
"<What>"
    "what" { NOGLOSS } %OBJ ACR CAPINIT PRON WH
"<I>"
    "i" { NOGLOSS } %SUBJ ACR PRON PERS NOM SG1
"<will>"
    "will" { FUT } %+FAUXV ACR V AUXMOD
"<make>"
    "make" { NOGLOSS } SVO %-FMAINV V INF
"<you>"
    "you" { NOGLOSS } %OBJ ACR PRON PERS ACC
"<think>"
    "think" { fikiri } HUM-V SVO %-FMAINV V INF
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { IMP-SG2 } V MW
"<,>"
    "," { , }
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<this>"
    "this" { h } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }
```

In the next phase we add appropriate tags to the verb, so that they can then be moved to appropriate places and converted to surface form (8).

```
(8)
"<What>"
    "what" { NOGLOSS } %OBJ ACR CAPINIT PRON WH
"<I>"
    "i" { NOGLOSS } %SUBJ ACR PRON PERS NOM SG1
"<will>"
    "will" { FUT } %+FAUXV ACR V AUXMOD SP-1
"<make>"
    "make" { NOGLOSS } SVO %-FMAINV V INF SP-SG1 SP-SG1 OP-SG2
"<you>"
    "you" { NOGLOSS } %OBJ ACR PRON PERS ACC
"<think>"
    "think" { fikiri } HUM-V SVO %-FMAINV V INF TAM-taka SP-SG1
    SP-SG1 REL-7 CAUS IMP-SG2 OP-SG2
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { IMP-SG2 } V MW
"<,>"
    "," { , }
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3 SP-SG1
"<this>"
```

```

    "this" { h } %PCOMPL-S PRON DEM SG DEM-9
"<.>"
    "." { . }

```

We see in (8) above that the verb *think* has several tags added for guiding the construction of the surface form. It seems a mess, and the tags are not in the order where they should be in the verb. Also, one tag was added twice. Confusion can be avoided so that, when moving tags as part of the verb, correct order is followed (9).

```

(9)
"<What>"
    "what" { NOGLOSS } %OBJ ACR CAPINIT PRON WH
"<I>"
    "i" { NOGLOSS } %SUBJ ACR PRON PERS NOM SG1
"<will>"
    "will" { SP-1+FUT } %+FAUXV ACR V AUXMOD
"<make>"
    "make" { SP-SG1+OP-SG2+NOGLOSS } SVO %-FMAINV V INF
"<you>"
    "you" { NOGLOSS } %OBJ ACR PRON PERS ACC
"<think>"
    "think" { SP-SG1+TAM-taka+REL-7+OP-SG2+fikir+Ish+eni } HUM-V
SVO %-FMAINV V INF
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { IMP-SG2 } V MW
"<,>"
    "," { , }
"<is>"
    "be" { SP-SG1+ni } MONOSLB %+FMAINV V PRES SG3
"<this>"
    "this" { h+DEM-9 } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

We see in (9) above that part of the tags moved to the verb are still tags, and others (those after the verb stem) are already converted to near-surface form. Now also the other tags will be converted to surface form (10).

```

(10)
"<What>"
    "what" { NOGLOSS } %OBJ ACR CAPINIT PRON WH
"<I>"
    "i" { NOGLOSS } %SUBJ ACR PRON PERS NOM SG1
"<will>"
    "will" { a+FUT } %+FAUXV ACR V AUXMOD
"<make>"
    "make" { ni+OP-SG2+NOGLOSS } SVO %-FMAINV V INF
"<you>"

```



```

    "you" { NOGLOSS } %OBJ ACR PRON PERS ACC
"<think>"
    "think" { ni+taka+cho+ku+fikir+Ish+eni } HUM-V SVO %-FMAINV
V INF
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { IMP-SG2 } V MW
"<,>"
    "," { , }
"<is>"
    "be" { ni } MONOSLB %+FMAINV V PRES SG3
"<this>"
    "this" { h+ii } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

We see that there are inflection tags also on words, which have no proper gloss, such as NOGLOSS, FUT and *ni*. These cause no harm, because they will be finally removed as non-proper glosses. The translation is in (11).

(11)
Nitakachokufikirisheni ni hii.

3 Swahili to English

When we translate to the other direction, the source sentence is *Nitakachokufikirisheni ni hii*. The sentence is first analysed in terms of morphology. In this implementation, glosses in English are included already in the morphological analyser. Note that the verb *fikiri* has three glosses. Also, the relative marker has two glosses. The object marker has two interpretations already at the morphological level. Therefore, the system produces two readings for the verb string (12).

```

(12)
"<<s>>"
    "<s>" { <s> }
"<*nitakachowafikirisheni>"
    "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-
REL { which , what } 2-PL2-OBJ OBJ { you } z [fikiri] { think ,
ponder , guess } SVO HUM-ACT CAUS:sh IMP-PL2 CAP
    "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-
REL { which , what } 2-PL3-OBJ OBJ { them } z [fikiri] { think ,
ponder , guess } SVO HUM-ACT CAUS:sh IMP-PL2 CAP
"<ni>"
    "ni" V V-BE INIT { it , he , she } { is }
    "ni" V V-BE INIT { they } { are }
    "ni" V V-BE NOSUBJ { is , are , am }
"<hii>"

```

```
"hii" PRON DEM :hV 4-PL { these }  
"hii" PRON DEM :hV 9-SG { this }  
"<.$>"  
  ".$" { .$ } **CLB  
"<<p>>"  
  "<p>" { <p> }  
"<<s>>"  
  "<s>" { <s> }
```

In order to disambiguate between all these alternatives, we must make a separate reading for each of them (13).

(13)

```
"<<s>>"  
  "<s>" { <s> }  
"<*nitakachowafikirisheni>"  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL2-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL2-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL2-OBJ OBJ { you } z [fikiri] { ponder } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL2-OBJ OBJ { you } z [fikiri] { ponder } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL2-OBJ OBJ { you } z [fikiri] { guess } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL2-OBJ OBJ { you } z [fikiri] { guess } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL3-OBJ OBJ { them } z [fikiri] { think } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { them } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL3-OBJ OBJ { them } z [fikiri] { ponder } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { them } z [fikiri] { ponder } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { which } 2-PL3-OBJ OBJ { them } z [fikiri] { guess } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP
```

```
"fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { them } z [fikiri] { guess } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
"<ni>"  
  "ni" V V-BE INIT { it } { is }  
  "ni" V V-BE INIT { he } { is }  
  "ni" V V-BE INIT { she } { is }  
  "ni" V V-BE INIT { they } { are }  
  "ni" V V-BE NOSUBJ { is }  
  "ni" V V-BE NOSUBJ { are }  
  "ni" V V-BE NOSUBJ { am }  
"<hii>"  
  "hii" PRON DEM :hV 4-PL { these }  
  "hii" PRON DEM :hV 9-SG { this }  
"<.$>"  
  ".$" { .$ } **CLB
```

Then we run the sentence through the disambiguation system (14).

```
(14)  
"<<s>>"  
  "<s>" { <s> }  
"<*nitakachowafikirisheni>"  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { ponder } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { guess } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP  
"<ni>"  
  "ni" V V-BE NOSUBJ { is }  
"<hii>"  
  "hii" PRON DEM :hV 9-SG { this }  
"<.$>"  
  ".$" { .$ } **CLB
```

In disambiguation process, most readings were removed. Only three readings of the verb were left, one for each gloss. There was no need to make full disambiguation, because the first of the three readings is considered the right one. The system is so constructed that if no rule applies, the first reading is selected (15).

```
(15)  
"<<s>>"  
  "<s>" { <s> }  
"<*nitakachowafikirisheni>"
```

```
"fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
<ni>  
  "ni" V V-BE INIT { is } @FMAINVintr-def  
<hii>  
  "hii" PRON DEM :hV 9-SG { this } @ADVL  
<.$>  
  ".$" { .$ } **CLB
```

Now we have only one reading for the verb. However, we have still to decide whether the subject prefix gloss and object prefix gloss are needed in this verb form. Therefore, we must add readings also for those cases, where the gloss is missing (16).

```
(16)  
<<s>>  
  <s> { <s> }  
<*nitakachowafikirisheni>  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ NO-OBJ-GLOSS z [fikiri] { think } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
  "fikirisha" V 1-SG1-SP VFIN NO-SP-GLOSS FUT:taka { will } 7-  
SG-REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { think } SVO  
HUM-ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
  "fikirisha" V 1-SG1-SP VFIN NO-SP-GLOSS FUT:taka { will } 7-  
SG-REL { what } 2-PL3-OBJ OBJ NO-OBJ-GLOSS z [fikiri] { think }  
SVO HUM-ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
<ni>  
  "ni" V V-BE INIT { is } @FMAINVintr-def  
<hii>  
  "hii" PRON DEM :hV 9-SG { this } @ADVL  
<.$>  
  ".$" { .$ } **CLB
```

Now we are in a situation, where we have four readings, and only one of them is correct. The gloss for the subject prefix must be present, and also the gloss for the object prefix must be present. Therefore, the first reading will be selected in this case (17).

```
(17)  
<<s>>  
  <s> { <s> }  
<*nitakachowafikirisheni>  
  "fikirisha" V 1-SG1-SP VFIN { *i } FUT:taka { will } 7-SG-  
REL { what } 2-PL3-OBJ OBJ { you } z [fikiri] { think } SVO HUM-  
ACT CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ>  
<ni>  
  "ni" V V-BE INIT { is } @FMAINVintr-def
```

```
"<hii>"  
  "hii" PRON DEM :hV 9-SG { this } @ADVL  
"<.$>"  
  ".$" { .$ } **CLB
```

The glosses must be rearranged into the order, where they are in target language. In order to do that, we will put the sentence on one line (18).

```
(18)  
( "<<s>>" ) ( { <s> } ) ( V 1-SG1-SP VFIN { *i } FUT:taka { will }  
7-SG-REL { what } 2-PL3-OBJ OBJ { you } z { think } SVO HUM-ACT  
CAUS:sh IMP-PL2 CAP @FMAINVtr-OBJ> ) ( V V-BE INIT { is }  
@FMAINVintr-def ) ( PRON DEM :hV 9-SG { this } @ADVL ) ( ".$" { .$  
} **CLB )
```

When the rearrangement rules are run, we get the result as in (19).

```
(19)  
( "<<s>>" ) ( { <s> } ) 7-SG-REL { what } :( V 1-SG1-SP VFIN { *i  
} FUT:taka { will } CAUS:sh :2-PL3-OBJ :OBJ { you } :z { think }  
SVO HUM-ACT :IMP-PL2 CAP @FMAINVtr-OBJ> ) ( V V-BE INIT { is }  
@FMAINVintr-def ) ( PRON DEM :hV 9-SG { this } @ADVL ) ( ".$" { .$  
} **CLB )
```

We still have two tags in the verb without gloss, CAUS:sh and IMP-PL2. We will add the glosses (20).

```
(20)  
( "<<s>>" )  
( { <s> } )  
7-SG-REL { *what } :( V 1-SG1-SP VFIN { *i } FUT:taka { will }  
CAUS:sh { make } V :2-PL3-OBJ :OBJ { you } :z { think } SVO HUM-  
ACT :IMP-PL2 { , each of you, } CAP @FMAINVtr-OBJ> )  
( V V-BE INIT { is } @FMAINVintr-def )  
( PRON DEM :hV 9-SG { this } @ADVL )  
( ".$" { .$ } **CLB )
```

When all tags are removed and only glosses are retained, we see that the words are in correct order (21).

```
(21)  
<s>  
*what  
*i  
will  
make  
you  
think
```

, each of you ,
is
this
.\$

The final translation is in (22).

(22)
What I will make you think, each of you, is this.

4 Swahili to Finnish

We make a third, even more challenging, test. We translate the sentence from Swahili to Finnish, although no translation system exists between these two languages. We solve the problem by translating first from Swahili to English, and then the translation result is translated to Finnish.

We have no need here to demonstrate the first phase here, because we already did that above. We concentrate on translation from English to Finnish. The analysed and disambiguated version of the sentence is in (23).

(23)

1	What	what		@OBJ %NH PRON WH
2	I	i	subj:>3	@SUBJ %NH PRON PERS NOM SG1
3	will	will	v-ch:>4	@+FAUXV %AUX V AUXMOD
4	make	make	subj:>12	@-FMAINV %VA V INF
5	you	you	subj:>6	@OBJ %NH PRON PERS ACC
6	think	think	obj:>4	@-FMAINV %VA V INF
7	,	,		
8	each	each	obj:>6	@OBJ %NH PRON SG
9	of	of	mod:>8	@<NOM-OF %N< PREP
10	you	you	pcomp:>9	@<P %NH PRON PERS ACC
11	,	,		
12	is	be	main:>0	@+FMAINV %VA V PRES SG3
13	this	this	comp:>12	@PCOMPL-S %NH PRON DEM SG
14	.	.		

We see that the result is as in (1) above. The analysis was done with the same analyser. Therefore, the result is the same. Differences start to be seen when we proceed in translation into Finnish. Here we perform the isolation of MWEs before we add Finnish glosses (24).

(24)

```
"<What>"
    "what" %OBJ CAPINIT PRON WH
"<I>"
    "i" %SUBJ PRON PERS NOM SG1
"<will>"
```

```
"will" %+FAUXV V AUXMOD
"<make>"
  "make" %-FMAINV V INF
"<you>"
  "you" %OBJ PRON PERS ACC
"<think>"
  "think" %-FMAINV V INF
"<,>"
  ", "
"<each_of_you>"
  "each_of_you" { jokainen N38 teistä , kukin teistä } %OBJ MW
PRON SG
"<,>"
  ", "
"<is>"
  "be" %+FMAINV V PRES SG3
"<this>"
  "this" %PCOMPL-S PRON DEM SG
"<.>"
  "."
```

The words each of you were combined as a MWE with two glosses. Then Finnish glosses are added to words (25).

(25)

```
"<What>"
  "what { mikä Np13 FRONT }" %OBJ CAPINIT PRON WH
"<I>"
  "i { minä Np5 , NOGLOSS , itse N8 FRONT } HUM OUT" %SUBJ
PRON PERS NOM SG1
"<will>"
  "will { NOGLOSS , aikoa V52-D , tulla V67 }" %+FAUXV V
AUXMOD
"<make>"
  "make { tehdä V71 FRONT O-PAR , tehdä V71 FRONT TRV O-TRA :2
, tehdä V71 FRONT O-ELA O-ACC :3 , tehdä V71 FRONT IO-ELA TRV :5 ,
tuottaa V53-C O-PAR , panna V67 V-3INF-ILL TRV , saada V63 V-3INF-
ILL TRV }" %-FMAINV V INF
"<you>"
  "you { sinä Np7 , te Np8 FRONT , NOGLOSS , itse N8 FRONT ,
sinun , teidän } HUM OUT" %OBJ PRON PERS ACC
"<think>"
  "think { ajatella V67-C O-PAR V-INF-PART } V-POS HUM-ACT" %-
FMAINV V INF
"<,>"
  ", "
"<each_of_you>"
  "each_of_you" { jokainen N38 teistä , kukin teistä } %OBJ MW
PRON SG
"<,>"
  ", "
```

```
"<is>"
    "be { olla V67b BE TRV-N V-4INF-TRA , olla V67b V-3INF-ILL ,
olla V67b V-3INF-INE , olla V67b BE O-PAR , eivät ole O-PAR V-
4INF-TRA :2 , eivät olleet O-PAR V-4INF-TRA :3 , emme :6 , emme
ole V-3INF-INE , emme olleet V-3INF-INE , ei ollut V-4INF-TRA , ei
ole O-PAR V-4INF-TRA , NOGLOSS , joka Np13 , jotka Np14 , tulla
V67 V-3INF-ILL } O-LOC1" %+FMAINV V PRES SG3
"<this>"
    "this { tämä Np1 FRONT , nämä Np2 FRONT }" %PCOMPL-S PRON
DEM SG
"<.>"
    "."
```

The sentence has two verbs, *make* and *be*, each with several glosses. Next, we put each gloss on separate line (26).

(26)

```
"<What>"
    "what" { mikä Np13 FRONT } %OBJ CAPINIT PRON WH
    "what" { PROP-CAND } %OBJ CAPINIT PRON WH
"<I>"
    "i" { minä Np5 } HUM OUT %SUBJ PRON PERS NOM SG1
    "i" { NOGLOSS } HUM OUT %SUBJ PRON PERS NOM SG1
    "i" { itse N8 FRONT } HUM OUT %SUBJ PRON PERS NOM SG1
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD
    "will" { aikoa V52-D } %+FAUXV V AUXMOD
    "will" { tulla V67 } %+FAUXV V AUXMOD
"<make>"
    "make" { tehdä V71 FRONT O-PAR } %-FMAINV V INF
    "make" { tehdä V71 FRONT TRV O-TRA :2 } %-FMAINV V INF
    "make" { tehdä V71 FRONT O-ELA O-ACC :3 } %-FMAINV V INF
    "make" { tehdä V71 FRONT IO-ELA TRV :5 } %-FMAINV V INF
    "make" { tuottaa V53-C O-PAR } %-FMAINV V INF
    "make" { panna V67 V-3INF-ILL TRV } %-FMAINV V INF
    "make" { saada V63 V-3INF-ILL TRV } %-FMAINV V INF
"<you>"
    "you" { sinä Np7 } HUM OUT %OBJ PRON PERS ACC
    "you" { te Np8 FRONT } HUM OUT %OBJ PRON PERS ACC
    "you" { NOGLOSS } HUM OUT %OBJ PRON PERS ACC
    "you" { itse N8 FRONT } HUM OUT %OBJ PRON PERS ACC
    "you" { sinun } HUM OUT %OBJ PRON PERS ACC
    "you" { teidän } HUM OUT %OBJ PRON PERS ACC
"<think>"
    "think" { ajatella V67-C O-PAR V-INF-PART } V-POS HUM-ACT %-
FMAINV V INF
"<,>"
    "," { , }
    "," { NOGLOSS }
"<each_of_you>"
    "each_of_you" { jokainen N38 teistä } %OBJ MW PRON SG
```



```
"each_of_you" { kukin teistä } %OBJ MW PRON SG
"<,>"
  "," { , }
  "," { NOGLOSS }
"<is>"
  "be" { olla V67b BE TRV-N V-4INF-TRA } O-LOC1 %+FMAINV V
PRES SG3
  "be" { olla V67b V-3INF-ILL } O-LOC1 %+FMAINV V PRES SG3
  "be" { olla V67b V-3INF-INE } O-LOC1 %+FMAINV V PRES SG3
  "be" { olla V67b BE O-PAR } O-LOC1 %+FMAINV V PRES SG3
  "be" { eivät ole O-PAR V-4INF-TRA :2 } O-LOC1 %+FMAINV V
PRES SG3
  "be" { eivät olleet O-PAR V-4INF-TRA :3 } O-LOC1 %+FMAINV V
PRES SG3
  "be" { emme :6 } O-LOC1 %+FMAINV V PRES SG3
  "be" { emme ole V-3INF-INE } O-LOC1 %+FMAINV V PRES SG3
  "be" { emme olleet V-3INF-INE } O-LOC1 %+FMAINV V PRES SG3
  "be" { ei ollut V-4INF-TRA } O-LOC1 %+FMAINV V PRES SG3
  "be" { ei ole O-PAR V-4INF-TRA } O-LOC1 %+FMAINV V PRES SG3
  "be" { NOGLOSS } O-LOC1 %+FMAINV V PRES SG3
  "be" { joka Np13 } O-LOC1 %+FMAINV V PRES SG3
  "be" { jotka Np14 } O-LOC1 %+FMAINV V PRES SG3
  "be" { tulla V67 V-3INF-ILL } O-LOC1 %+FMAINV V PRES SG3
"<this>"
  "this" { tämä Np1 FRONT } %PCOMPL-S PRON DEM SG
  "this" { nämä Np2 FRONT } %PCOMPL-S PRON DEM SG
"<.>"
  "." { . }
```

This ambiguous sentence will be disambiguated (27).

(27)

```
"<What>"
  "what" { mikä Np13 FRONT } %OBJ CAPINIT PRON WH
"<I>"
  "i" { minä Np5 } HUM OUT %SUBJ PRON PERS NOM SG1
"<will>"
  "will" { NOGLOSS } %+FAUXV V AUXMOD
"<make>"
  "make" { panna V67 V-3INF-ILL TRV } %-FMAINV V INF
"<you>"
  "you" { te Np8 FRONT } HUM OUT %OBJ PRON PERS ACC
"<think>"
  "think" { ajatella V67-C O-PAR V-INF-PART } V-POS HUM-ACT %-
FMAINV V INF
"<,>"
  "," { , }
"<each_of_you>"
  "each_of_you" { jokainen N38 teistä } %OBJ MW PRON SG
"<,>"
  "," { , }
```

```
"<is>"
    "be" { olla V67b V-3INF-ILL } O-LOC1 %+FMAINV V PRES SG3
"<this>"
    "this" { tämä Np1 FRONT } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }
```

Now we have one reading for each word. Next, we add inflection tags (28).

```
(28)
"<What>"
    "what" { mikä Np13 FRONT } %OBJ CAPINIT PRON WH SG PAR
"<I>"
    "i" { minä Np5 } %SUBJ HUM OUT PRON PERS SG1
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD SG1 PRES
"<make>"
    "make" { panna V67 } %-FMAINV V-3INF-ILL TRV V INF SG1 PRES
"<you>"
    "you" { te Np8 FRONT } %OBJ HUM OUT PRON PERS PL2 ACC
"<think>"
    "think" { ajatella V67-C } %-FMAINV O-PAR V-INF-PART V-POS
HUM-ACT V INF SG1 3INF-ILL
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { jokainen N38 teistä } %OBJ MW PRON SG ACC
"<,>"
    "," { , }
"<is>"
    "be" { olla V67b } %+FMAINV V-3INF-ILL O-LOC1 V PRES SG3
"<this>"
    "this" { tämä Np1 FRONT } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }
```

When the inflection tags are moved to the appropriate places and converted to surface form, we get the result as in (29).

```
(29)
"<What>"
    "what" { mi:kä+ta :Np13 FRONT } %OBJ CAPINIT PRON WH SG PAR
"<I>"
    "i" { min:ä :Np5 } %SUBJ HUM OUT PRON PERS SG NOM
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD SG1 PRES
"<make>"
    "make" { pan:na+en :V67 } %-FMAINV V-3INF-ILL TRV V SG1 PRES
"<you>"
```

```

    "you" { t:e+eidät :Np8 FRONT } %OBJ HUM OUT PRON PERS PL2
ACC
"<think>"
    "think" { ajatt%el:la+emaan :V67-C } %-FMAINV O-PAR V-INF-
PART V-POS HUM-ACT V SG1 3INF-ILL
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { jokai:nen+sen :N38 teistä } %OBJ MW PRON SG
ACC
"<,>"
    "," { , }
"<is>"
    "be" { o:lla+n :V67b } %+FMAINV V-3INF-ILL O-LOC1 V PRES SG3
"<this>"
    "this" { tä:mä :Np1 FRONT } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

The glosses above are not final glosses. They consist of three parts. The first part is the non-inflecting part of the stem. The middle part is the last part of the base form. The last part is the inflected ending of the word. The middle part will ultimately disappear. This format is just to help to see the full base form of the word. The glosses are clearer in (30).

```

(30)
"<What>"
    "what" { mi+tä } %OBJ CAPINIT PRON WH SG PAR
"<I>"
    "i" { min:ä } %SUBJ HUM OUT PRON PERS SG NOM
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD SG1 PRES
"<make>"
    "make" { pan+en } %-FMAINV V-3INF-ILL TRV V SG1 PRES
"<you>"
    "you" { t+eidät } %OBJ HUM OUT PRON PERS PL2 ACC
"<think>"
    "think" { ajatt%el+emaan } %-FMAINV O-PAR V-INF-PART V-POS
HUM-ACT V SG1 3INF-ILL
"<,>"
    "," { , }
"<each_of_you>"
    "each_of_you" { jokai+sen teistä } %OBJ MW PRON SG ACC
"<,>"
    "," { , }
"<is>"
    "be" { o+n } %+FMAINV V-3INF-ILL O-LOC1 V PRES SG3
"<this>"
    "this" { tä:mä } %PCOMPL-S PRON DEM SG
"<.>"
    "." { . }

```

We see that also the concordance was controlled in this phase. Because there is no need to rearrange words, we can produce the final translation (31).

(31)

Mitä minä panen teidät ajattelemaan, jokaisen teistä, on tämä.

5 Discussion

We have seen that the quite complex Swahili verb form was translated into English. The translation is not elegant, but it is not possible to translate it otherwise without losing information. Also, the translation from English to Swahili succeeded faultlessly.

The translation from Swahili into Finnish was done in two steps, first from Swahili into English, and then from English into Finnish. Even this test succeeded without fault.

I also tested how Google Translate succeeds in this test. Below are results.

Swahili to English

Source sentence: Nitakachokufikirisheni ni hii.

SALAMA: *What I will make you think, each of you, is this.*

GT: *Here is what I will bring you.*

English to Swahili

Source sentence: What I will make you think, each of you, is this.

SALAMA: *Nitakachokufikirisheni ni hii.*

GT: *Kile nitakachokufanya ufikirie, kila mmoja wako, ni hii.*

Swahili to Finnish

Source sentence: Nitakachokufikirisheni ni hii.

SALAMA: *Mitä minä panen teidät ajattelemaan, jokaisen teistä, on tämä.*

GT: *Tässä on mitä minä tuon sinulle*

The comparison between these two translation systems shows how difficult, if not impossible, it is to translate complex structures using such approaches to translation as Google does. It is not possible to understand how GT manages to interpret the main verb *fikiri* in the way it does.

Also, in the translation from English to Swahili, GT does not even try to construct the correct Swahili structure. However, the translation, with another structure, is quite good, although with a minor grammatical error (*wako > wenu*).

The tests demonstrate that using rule-based approaches to translation it is hard to find such a translation task, which cannot be translated correctly.