Technical Reports in Language Technology
Report No 22, 2017
http://www.njas.helsinki.fi/salama

Machine translation through interlingua

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

Abstract

The report demonstrates how English can be used as interlingua in translating between two structurally very different languages, such as Swahili and Finnish. It shows also that some problems of English language usage can be bypassed in this process. Consequently, the translation quality from Swahili through English to Finnish is better than the translation quality from English to Finnish.

Keywords: *interlingua, machine translation, multiword expressions, disambiguation.*

1 Introduction

For a long time there has been the idea that interlingua could serve as a kind of black box when translating from one language to another. If there would be such a 'switchboard', through which translation could be made, a lot of work could be saved in constructing translation applications between languages. In spite of several attempts to devise an interlingua, none of the attempts has gained wide acceptance. Esperanto or other attempts based on Romance languages are not widely enough known in the world, and it is counterintuitive to learn one more language just for the sake of machine translation. The fact is that English has established itself as world language, as a kind of global lingua franca. It is learned in schools around the world, and much of business in the world is carried out in this language.

English is far from ideal as interlingua. In a way it is a 'worn-out' language, which has lost a number of such linguistic features, which would help in machine translation. The ambiguity of English words is a constant problem. Ambiguity extends to such basic levels as part-of-speech, not to talk of more fine-grained semantic features. Another annoying feature is that English native speakers tend to omit relative pronouns and some conjunctions that initiate a subordinate clause. The human being is usually able to understand what the speaker means, but for a computer it is often a nightmare. It is a known fact that if people would be more systematic and explicit in using language, it would be much easier to implement machine translation systems. But people are what they are, prone to 'simplify' structures and make all sorts of errors in language use. But we have to cope with the situation.

In this report I will show that it is possible to 'correct' English when using it as interlingua. It is possible to produce such English, which makes explicit such features, which in current English are often hidden. Below I will demonstrate this using Swahili as source language and Finnish as target language. Both Swahili and Finnish have complex linguistic structure. They are very different from English and also from each other. The demonstration shows that it is easier to translate from Swahili via English to Finnish than to translate the same from English to Finnish.

2 Translation process from Swahili to Finnish

Problems of translation are demonstrated below using a single sentence picked from news media. The sentence contains problems described in introduction. It also has a passive structure, which cannot be used in Finnish, but can be used in Swahili.

Original sentence in English news media:

(1) This means a lot of important research is not seen or read by scientists and researchers.

When more clearly written, the sentence has the form as in (2).

(2) This means that a lot of important research is not seen or read by scientists and researchers.

It is possible to write the sentence also into form, where the sections of the sentence are still more clearly displayed (3).

(3) This means that there is a lot of important research which is not seen or read by scientists and researchers.

2.1 Translation from Swahili to English

When the sentence is translated into Swahili or Finnish, it follows the format as in (3). Relative pronouns and conjunctions initiating a subordinate clause must be made explicit. Let us see phase by phase how translation from Swahili to Finnish proceeds. Sentence (3) translated to Swahili is as in (4).

(4) Hii inamaanisha kwamba kuna kwa wingi utafiti muhimu, ambao hauonekani wala hausomwi na wanasayansi wala watafiti.

When the sentence is analysed and disambiguated, we get the result as in (5).

(5)

```
"<*hii>"
   "hii" PRON DEM :hV 9-SG { this } INITCAP @SUBJ
"<inamaanisha>"
```

```
      "maanisha" V 9-SG-SP VFIN NO-SP-GLOSS PR:na z [maanika] {
mean } SVO @FMAINVtr-OBJ>
   "<kwamba>"
      "kwamba" CONJ { that } **CLB @CS
   "<kuna>"
      "kuna" V 17-SG-SP VFIN C:na { there is } @FMAINVintr
   "<kwa_wingi>"
      "kwa_wingi" ADV { a lot } @ADVL
   "<utafiti>"
      "utafiti" N 11/10-SG { the } { research } AR @SUBJ
   "<muhimu>"
      "muhimu" ADJ A-UNINFL { important } @<NADJ
   "<,>"
      "," COMMA { , }
   "<ambao>"
      "amba" PRON **SUBJ-CLB AMBA-REL REL 11-SG { which } @SUBJ
   "<hauonekani>"
      "onekana" V NEG-a 11-SG-SP VFIN NO-SP-GLOSS z [ona] { be
seen } SV PREFR REC @FMAINVintr
   "<wala>"
      "wala" ADV { nor } AR @ADVL
   "<hausomwi>"
      "somwa" V NEG-a 11-SG-SP VFIN { it } z [soma] { read } SVO
SVOO PASS @FMAINVtr-OBJ>
   "<na>"
      "na" AG-PART { by } @ADVL
   "<wanasayansi>"
      "mwanasayansi" N 1/2-PL HUM { the } { scientist } ENG @AG
   "<wala>"
      "wala" ADV { nor } AR @ADVL
   "<watafiti>"
      "mtafiti" N 1/2-PL HUM { the } { researcher } @OBJ
   "<.$>"
      ".$" { .$ } **CLB
```

After several phases of processing we get the form as in (6). The original text is removed and the text of target language and grammatical information is retained. Also the correct word order is implemented. Note that the colon ':' is inserted in front of various tags to prevent further rule application.

(6)

```
   ( PRON DEM :hV 9-SG { *this } INITCAP @SUBJ )
   ( V 9-SG-SP VFIN NO-SP-GLOSS PR:na :z { :means } SVO @FMAINVtr-
OBJ> )
   ( CONJ { that } **CLB @CS )
   ( V 17-SG-SP VFIN C:na { there is } @FMAINVintr )
   ( ADV { a lot of } @ADVL )
   :( ADJ A-UNINFL { important } @<NADJ )
   :( N 11/10-SG { research } @SUBJ )
   ( COMMA { , } )
```

```
   ( PRON **SUBJ-CLB AMBA-REL REL 11-SG { which } @SUBJ )
   ( V NEG-a 11-SG-SP VFIN NO-SP-GLOSS :z { :does not :be :seen }
SV PREFR REC @FMAINVintr )
   ( ADV { nor } @ADVL )
   ( V NEG-a 11-SG-SP VFIN { it } :z { :is } { not } { :read } SVO
SVOO PASS @FMAINVtr-OBJ> )
   ( AG-PART { by } @ADVL )
   ( N 1/2-PL HUM { the } { scientists } @AG )
   ( ADV { nor } @ADVL )
   ( N 1/2-PL HUM { the } { researchers } @OBJ )
   ( ".$" { .$ } **CLB )
```

After removing unnecessary tags we get the translation to English as in (7).

(7)
This means that there is a lot of important research, which does not be seen nor it is not read by the scientists nor the researchers.

The text in (7) is not yet final translation. It still resembles too much the original Swahili text and has traces of Swahili grammar. This can be corrected in a controlled way through post processing, as in (8).

(8)
This means that there is a lot of important research, which is neither seen nor read by the scientists or the researchers.

The translation in (8) produces definite articles, although the original English text does not have them. It is very difficult to produce correct level of definitness in English, if the source language does not mark definitness systematically. But we will see that this does not matter, because the final target language (Finnish) does not have articles either.

2.2 Translation from interlingua (English) to Finnish

When we pass the translated text in (8) to the Finnish translator, we get initially the result as in (9). The translator firs analyses the English text.

(9)

```
"<*this>"
   "this" %SUBJ PRON DEM SG CAPINIT
"<means>"
   "mean" %+FMAINV V PRES SG3
"<that>"
   "that" %CS CS
"<there>"
   "there" %F-SUBJ <Ex> ADV
"<is>"
   "be" %+FMAINV V PRES SG3
"<a lot of>"
```

```
      "a lot of" %DN> DET
   "<important>"
      "important" %A> A ABS
   "<research>"
      "research" %PCOMPL-S N SG NOM
   "<,>"
      ","
   "<which>"
      "which" %SUBJ <Rel> PRON WH NOM
   "<is>"
      "be" %+FAUXV V PRES SG3
   "<neither>"
      "neither" %ADVL ADV
   "<seen>"
      "see" %-FMAINV EN
   "<nor>"
      "nor" %CC CC
   "<read>"
      "read" %-FMAINV EN-AG
   "<by>"
      "by" %ADVL PREP
   "<scientists>"
      "scientist" %<P N PL NOM
   "<or>"
      "or" %CC CC
   "<researchers>"
      "researcher" %<P N PL NOM
   "<.>"
      "."
```

When we add the Finnish glosses and grammatical tags for directing translation, we get the result as in (10).

(10)

```
   "<*this>"
      "this" { tämä Np1 FRONT , nämä Np2 FRONT } %SUBJ PRON DEM SG
CAPINIT
   "<means>"
      "mean" { tarkoittaa V53-C O-PAR , merkitä V69 FRONT O-PAR }
%+FMAINV V PRES SG3
   "<that>"
      "that" { COMMA että , COMMA jotta , COMMA joka Np13 ,
NOGLOSS M-GEN , COMMA jolloin } %CS CS
   "<there>"
      "there" { siinä , siellä , sinne , NOGLOSS , tuolla , tuonne
} LOC %F-SUBJ <Ex> ADV
   "<is>"
      "be" { olla V67b BE , eivät ole , ei ole , NOGLOSS , joka
Np13 , jotka Np14 } %+FMAINV V PRES SG3
   "<a lot of>"
      "a lot of" { paljon } M-PAR %DN> DET
```

```
   "<important>"
      "important" { tärkeä N15 FRONT , tärkeämpi N16-H FRONT , N51
FRONT , merkittävä N10 FRONT } %A> A ABS
   "<research>"
      "research" { tutkimus N39 } %PCOMPL-S N SG NOM
   "<,>"
      "," { COMMA , NOGLOSS }
   "<which>"
      "which" { COMMA joka Np13 } %SUBJ <Rel> PRON WH NOM
   "<is>"
      "be" { olla V67b BE , eivät ole , ei ole , NOGLOSS , joka
Np13 , jotka Np14 } %+FAUXV V PRES SG3
   "<neither>"
      "neither" { ei , eikä } %ADVL ADV
   "<seen>"
      "see" { nähdä V71 FRONT O-ACC , katso O-PAR , on nähtävissä
M-INE } %-FMAINV EN
   "<nor>"
      "nor" { eikä , eivätkä } %CC CC
   "<read>"
      "read" { lukea V58-D O-PAR } %-FMAINV EN-AG
   "<by>"
      "by" { kanssa M-GEN , mennessä M-ILL , NOGLOSS M-ADE ,
NOGLOSS AG-PART , NOGLOSS M-INS , jonka , jotka , avulla M-GEN }
POST %ADVL PREP
   "<scientists>"
      "scientist" { tiedemies N42 FRONT , tutkija N12 , tieteilijä
N12 FRONT } %<P N PL NOM
   "<or>"
      "or" { tai , vai } %CC CC
   "<researchers>"
      "researcher" { tutkija N12 } %<P N PL NOM
   "<.>"
      "." { . }
```

We see that many words have several interpretations, which calls for disambiguation. A more readable version of the above is in (11).

(11)

```
   "<*this>"
      "this" { tämä Np1 FRONT }  %SUBJ PRON DEM SG CAPINIT
      "this" { nämä Np2 FRONT }  %SUBJ PRON DEM SG CAPINIT
   "<means>"
      "mean" { tarkoittaa V53-C O-PAR }  %+FMAINV V PRES SG3
      "mean" { merkitä V69 FRONT O-PAR }  %+FMAINV V PRES SG3
   "<that>"
      "that" { , että } %CS CS
      "that" { , jotta } %CS CS
      "that" { , joka Np13 } %CS CS
      "that" { NOGLOSS M-GEN } %CS CS
      "that" { , jolloin } %CS CS
```

```
"<there>"
   "there" { siinä } LOC %F-SUBJ <Ex> ADV
   "there" { siellä } LOC %F-SUBJ <Ex> ADV
   "there" { sinne } LOC %F-SUBJ <Ex> ADV
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
   "there" { tuolla } LOC %F-SUBJ <Ex> ADV
   "there" { tuonne } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { olla V67b BE } %+FMAINV V PRES SG3
   "be" { eivät ole } %+FMAINV V PRES SG3
   "be" { ei ole } %+FMAINV V PRES SG3
   "be" { NOGLOSS } %+FMAINV V PRES SG3
   "be" { joka Np13 } %+FMAINV V PRES SG3
   "be" { jotka Np14 } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärkeä N15 FRONT } %A> A ABS
   "important" { tärkeämpi N16-H FRONT } %A> A ABS
   "important" { N51 FRONT } %A> A ABS
   "important" { merkittävä N10 FRONT } %A> A ABS
"<research>"
   "research" { tutkimus N39 } %PCOMPL-S N SG NOM
"<,>"
   "," { , }
   "," { NOGLOSS }
"<which>"
   "which" { , joka Np13 } %SUBJ <Rel> PRON WH NOM
"<is>"
   "be" { olla V67b BE } %+FAUXV V PRES SG3
   "be" { eivät ole } %+FAUXV V PRES SG3
   "be" { ei ole } %+FAUXV V PRES SG3
   "be" { NOGLOSS } %+FAUXV V PRES SG3
   "be" { joka Np13 } %+FAUXV V PRES SG3
   "be" { jotka Np14 } %+FAUXV V PRES SG3
"<neither>"
   "neither" { ei } %ADVL ADV
   "neither" { eivät } %ADVL ADV

"<seen>"
   "see" { nähdä V71 FRONT O-ACC } %-FMAINV EN
   "see" { katso O-PAR } %-FMAINV EN
   "see" { on nähtävissä M-INE } %-FMAINV EN
"<nor>"
   "nor" { eikä } %CC CC
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lukea V58-D O-PAR } %-FMAINV EN-AG
"<by>"
   "by" { kanssa M-GEN } POST %ADVL PREP
   "by" { mennessä M-ILL } POST %ADVL PREP
```

```
      "by" { NOGLOSS M-ADE } POST %ADVL PREP
      "by" { NOGLOSS AG-PART } POST %ADVL PREP
      "by" { NOGLOSS M-INS } POST %ADVL PREP
      "by" { jonka } POST %ADVL PREP
      "by" { jotka } POST %ADVL PREP
      "by" { avulla M-GEN } POST %ADVL PREP
   "<scientists>"
      "scientist" { tiedemies N42 FRONT } %<P N PL NOM
      "scientist" { tutkija N12 } %<P N PL NOM
      "scientist" { tieteilijä N12 FRONT } %<P N PL NOM
   "<or>"
      "or" { tai }  %CC CC
      "or" { vai }  %CC CC
   "<researchers>"
      "researcher" { tutkija N12 } %<P N PL NOM
   "<.>"
      "." { . }
```

As can be seen in (11), words may have several semantic interpretations when compared with target language. The major difference between English and Finnish is that English uses prepositions for expressing semantic relations and Finnish uses most often cases. Mapping between these two systems is very complex. Below we will see how this can be done.

After semantic disambiguation, the result is as in (12).

(12)

```
   "<*this>"
      "this" { tämä Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
   "<means>"
      "mean" { tarkoittaa V53-C } O-PAR %+FMAINV V PRES SG3
   "<that>"
      "that" { , että } %CS CS
   "<there>"
      "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
   "<is>"
      "be" { olla V67b BE } %+FMAINV V PRES SG3
   "<a lot of>"
      "a lot of" { paljon } M-PAR %DN> DET
   "<important>"
      "important" { tärkeä N15 FRONT } %A> A ABS
   "<research>"
      "research" { tutkimus N39 } %PCOMPL-S N SG NOM
   "<,>"
      "," { , }
   "<which>"
      "which" { , joka Np13 } %SUBJ <Rel> PRON WH NOM
   "<is>"
      "be" { NOGLOSS } %+FAUXV V PRES SG3
   "<neither>"
      "neither" { eivät } %ADVL ADV
```

```
"<seen>"
   "see" { nähdä V71 FRONT } O-ACC %-FMAINV EN
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lukea V58-D } O-PAR %-FMAINV EN-AG
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemies N42 FRONT } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkija N12 } %<P N PL NOM
"<.>"
   "." { . }
```

In this stage, the linguistic tags added to readings are context-insensitive tags. For example, each verb is assigned a tag according to what case it requires for the object. For instance, "see" gets the tag O-ACC meaning that the object should be in accusative case. The verb "read" has a tag O-PAR indicating that the object should be in partitive case. If the verb is intransitive, neither of these two tags is added. We should note, however, that the tags of transitivity are not always safe tags. Many transitive verbs may have the object in accusative or partitive, depending on the case.

Each nominal and verb also has a tag for defining the inflection class. There is a total of 76 basic classes for inflection. In addition, there are several inflection classes for a limited number of words. Inflection classes for nouns are marked with 'N' followed by a numerical code from 1 to 51 (e.g. N42). Verbs are marked with 'V' followed by a numerical code from 52 to 74. There are also verbs that do not fall into the established classes (e.g. V67b).

To make inflection still more complicated, there is a total of 13 classes for gradation. Words that are subject to gradation have a weak grade in some forms and a strong grade in others. Grades are marked in notation with capital letters from A to M (e.g. V58-D).

The inflection also depends on the vowel quality of the stem. In this system, back vowel words are the default and are without marking. Front vowel words have the tag FRONT.

3 Adding inflection tags

Tags for inflection are added to words in a strict order. The very first thing is to make sure that each word has the appropriate tag for number (singular or plural). The possible tags are SG, PL, SG1, SG2, SG3, PL1, PL2 and PL3. The analysis result from English contains some tags, but many more need to be added, and some others need to be changed. In (13) tags that are added are prefixed with '@'.

(13)

```
"<*this>"
   "this" { tämä Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
```

```
"<means>"
   "mean" { tarkoittaa V53-C } O-PAR %+FMAINV V PRES SG3
"<that>"
   "that" { , että } %CS CS
"<there>"
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { olla V67b BE } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärkeä N15 FRONT } %A> A ABS
"<research>"
   "research" { tutkimus N39 } %PCOMPL-S N SG NOM
"<,>"
   "," { , }
"<which>"
   "which" { , joka Np13 } %SUBJ <Rel> PRON WH NOM @SG
"<is>"
   "be" { NOGLOSS } %+FAUXV V PRES SG3
"<neither>"
   "neither" { eivät } %ADVL ADV
"<seen>"
   "see" { nähdä V71 FRONT } O-ACC %-FMAINV EN @PL
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lukea V58-D } O-PAR %-FMAINV EN-AG @PL
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemies N42 FRONT } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkija N12 } %<P N PL NOM
"<.>"
   "." { . }
```

The tags for singular and plural are so called primary level tags, as are also the tags for inflecting the main constituents of the sentence. The primary level constituents include the subject, the main verb, the object, and other types of modifiers of the verb. These tags are displayed in (14).

(14)

```
"<*this>"
   "this" { tämä Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
"<means>"
   "mean" { tarkoittaa V53-C } O-PAR %+FMAINV V PRES SG3 SG
"<that>"
```

```
   "that" { , että } %CS CS
"<there>"
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { olla V67b BE } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärkeä N15 FRONT } %A> A ABS
"<research>"
   "research" { tutkimus N39 } %PCOMPL-S N SG NOM @PAR
"<,>"
   "," { , }
"<which>"
   "which" { , joka Np13 } %SUBJ <Rel> PRON WH NOM SG @PAR
"<is>"
   "be" { NOGLOSS } %+FAUXV V PRES SG3 @PASS-PRES
"<neither>"
   "neither" { eivät } %ADVL ADV
"<seen>"
   "see" { nähdä V71 FRONT } O-ACC %-FMAINV EN PL @NEG-PRES
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lukea V58-D } O-PAR %-FMAINV EN-AG PL @NEG-PRES
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemies N42 FRONT } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkija N12 } %<P N PL NOM
"<.>"
   "." { . }
```

Tags are added on the basis of information available in context. For example, the word "research" gets the tag @PAR, because the determiner "a lot of" has the tag M-PAR, indicating that the following noun should be in partitive. The inflection tags for the verbs "see" and "read" are much more difficult to define, because English has a passive structure, and Finnish does not have a corresponding structure at all. Finnish has only a neutro-passive and not a passive constructed with an agent. Therefore, the verbs should be translated with active forms. Also the word order must be fundamentally changed.

When the first level inflection tags are added, we add the second level tag. Such tags include noun modifiers, such as adjectives, numerals and pronouns. This is demonstrated in (15).

(15)

```
"<*this>"
   "this" { tämä Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
```

```
"<means>"
   "mean" { tarkoittaa V53-C } O-PAR %+FMAINV V PRES SG3 SG
"<that>"
   "that" { , että } %CS CS
"<there>"
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { olla V67b BE } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärkeä N15 FRONT } %A> A ABS SG NOM @PAR
"<research>"
   "research" { tutkimus N39 } %PCOMPL-S N SG NOM PAR
"<,>"
   "," { , }
"<which>"
   "which" { , joka Np13 } %SUBJ <Rel> PRON WH SG PAR
"<is>"
   "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
"<neither>"
   "neither" { eivät } %ADVL ADV
"<seen>"
   "see" { nähdä V71 FRONT } O-ACC %-FMAINV EN PL NEG-PRES
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lukea V58-D } O-PAR %-FMAINV EN-AG PL NEG-PRES
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemies N42 FRONT } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkija N12 } %<P N PL NOM NOM
"<.>"
   "." { . }
```

In this example there is only one secondary tag. The adjective "important" gets the tag @PAR, because it is the modifier of the following noun, which already was assigned the tag PAR in earlier phase. The reader may be confused be seeing that a tag earlier prefixed with '@' has later lost the prefix. This is a purely technical solution. The prefix '@' is inserted to tags, because it makes the tag 'immune' in the sense that no other tag with the same prefix can be inserted to the reading. This facility helps in ordering the rules so that secure rules have preference, and less secure rules will apply only in the case that no secure rule has applied. After each section of rule application the prefix is removed, so that new rules in the subsequent phase can be applied.

4 Converting inflection tags to surface form

Each inflection tag is converted to surface form on the basis of the information attached to glosses. Nouns and adjectives contain information on the case class, gradation class, and on its back/front affiliation. Verbs have information on verb class, gradation class, and on its affiliation in back/front dichotomy. All information needed for producing the surface form of the two inflection tags (SG/PL plus case tag or verb tag).

Before we start this process, we have to mark the point, where the word stem ends (16).

(16)

```
"<*this>"
   "this" { tä:mä Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
"<means>"
   "mean" { tarkoitt:aa V53-C } O-PAR %+FMAINV V PRES SG3 SG
"<that>"
   "that" { , että } %CS CS
"<there>"
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { o:lla V67b BE } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärke:ä N15 FRONT } %A> A ABS SG PAR
"<research>"
   "research" { tutkimu:s N39 } %PCOMPL-S N SG PAR
"<,>"
   "," { , }
"<which>"
   "which" { , jo:ka Np13 } %SUBJ <Rel> PRON WH SG PAR
"<is>"
   "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
"<neither>"
   "neither" { eivät } %ADVL ADV
"<seen>"
   "see" { nä:hdä V71 FRONT } O-ACC %-FMAINV EN PL NEG-PRES
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { luk:ea V58-D } O-PAR %-FMAINV EN-AG PL NEG-PRES
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemie:s N42 FRONT } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkij:a N12 } %<P N PL NOM NOM
"<.>"
   "." { . }
```

We see that part of the end section of the word is not considered the stem proper. The number of letters thus cut off from the end varies between zero and five (16). Next we convert the inflection tag into surface form (17).

```
(17)
   "<*this>"
      "this" { tä:mä :Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
   "<means>"
      "mean" { tarkoitt:aa :V53-C } O-PAR %+FMAINV V PRES SG3 +aa
   "<that>"
      "that" { , että } %CS CS
   "<there>"
      "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
   "<is>"
      "be" { o:lla :V67b } %+FMAINV V PRES SG3 +n
   "<a lot of>"
      "a lot of" { paljon } M-PAR %DN> DET
   "<important>"
      "important" { tärke:ä :N15 FRONT } %A> A ABS SG PAR +aa
   "<research>"
      "research" { tutkimu:s :N39 } %PCOMPL-S N SG PAR +sta
   "<,>"
      "," { , }
   "<which>"
      "which" { , jo:ka :Np13 } %SUBJ <Rel> PRON WH SG PAR +ta
   "<is>"
      "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
   "<neither>"
      "neither" { eivät } %ADVL ADV
   "<seen>"
      "see" { nä:hdä :V71 FRONT } O-ACC %-FMAINV PL NEG-PRES +e
   "<nor>"
      "nor" { eivätkä } %CC CC
   "<read>"
      "read" { luk:ea :V58-D } O-PAR %-FMAINV EN-AG PL NEG-PRES +e
   "<by>"
      "by" { NOGLOSS AG-PART } POST %ADVL PREP
   "<scientists>"
      "scientist" { tiedemie:s :N42 FRONT } %<P N PL NOM +het
   "<or>"
      "or" { tai } %CC CC
   "<researchers>"
      "researcher" { tutkij:a :N12 } %<P N PL NOM +at NOM +at
   "<.>"
      "." { . }
```

Each inflection tag in (17) is converted into surface form and prefixed by '+'. In the next phase we move the surface form to the end of the gloss (18).

```
(18)
   "<*this>"
      "this" { tä:mä :Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
   "<means>"
      "mean" { tarkoitt:aa+aa :V53-C } O-PAR %+FMAINV V PRES SG3
   "<that>"
      "that" { , että } %CS CS
   "<there>"
      "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
   "<is>"
      "be" { o:lla+n :V67b } %+FMAINV V PRES SG3
   "<a lot of>"
      "a lot of" { paljon } M-PAR %DN> DET
   "<important>"
      "important" { tärke:ä+aa :N15 FRONT } %A> A ABS SG PAR
   "<research>"
      "research" { tutkimu:s+sta :N39 } %PCOMPL-S N SG PAR
   "<,>"
      "," { , }
   "<which>"
      "which" { , jo:ka+ta :Np13 } %SUBJ <Rel> PRON WH NOM SG NOM
   "<is>"
      "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
   "<neither>"
      "neither" { eivät } %ADVL ADV
   "<seen>"
      "see" { nä:hdä+e :V71 FRONT } O-ACC %-FMAINV PL NEG-PRES
   "<nor>"
      "nor" { eivätkä } %CC CC
   "<read>"
      "read" { luk:ea+e :V58-D } O-PAR %-FMAINV EN-AG PL NEG-PRES
   "<by>"
      "by" { NOGLOSS AG-PART } POST %ADVL PREP
   "<scientists>"
      "scientist" { tiedemie:s+het :N42 FRONT } %<P N PL NOM
   "<or>"
      "or" { tai } %CC CC
   "<researchers>"
      "researcher" { tutkij:a+at :N12 } %<P N PL NOM
   "<.>"
      "." { . }
```

We see in (18) that the words now have the entire gloss plus the inflected ending. Although the section between ':' and '+' will be finally removed, the section is still needed, because it helps to define the precise surface form of the word.

5 Handling gradation and front/back concordance

When the glosses now have the inflection suffixes attached to them, we can handle gradation and front/back affiliation. In (19) we have implemented the strong/weak grade

variation. It turns out that the example sentence has only one instance of this variation ("read"), and the other instance ("mean") has strong grade, which is considered default.

(19)

```
   "<*this>"
      "this" { tä:mä :Np1 FRONT } %SUBJ PRON DEM SG CAPINIT
   "<means>"
      "mean" { tarkoitt:aa+aa :V53-C } O-PAR %+FMAINV V PRES SG3
+aa SG
   "<that>"
      "that" { , että } %CS CS
   "<there>"
      "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
   "<is>"
      "be" { o:lla+n :V67b } %+FMAINV V PRES SG3
   "<a lot of>"
      "a lot of" { paljon } M-PAR %DN> DET
   "<important>"
      "important" { tärke:ä+aa :N15 FRONT } %A> A ABS SG PAR
   "<research>"
      "research" { tutkimu:s+sta :N39 } %PCOMPL-S N SG PAR
   "<,>"
      "," { , }
   "<which>"
      "which" { , jo:ka :Np13 } %SUBJ <Rel> PRON WH NOM SG NOM
   "<is>"
      "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
   "<neither>"
      "neither" { eivät } %ADVL ADV
   "<seen>"
      "see" { nä:hdä+e :V71 FRONT } O-ACC %-FMAINV PL NEG-PRES
   "<nor>"
      "nor" { eivätkä } %CC CC
   "<read>"
      "read" { lu%:ea+e :V58-D } O-PAR %-FMAINV EN-AG PL NEG-PRES
   "<by>"
      "by" { NOGLOSS AG-PART } POST %ADVL PREP
   "<scientists>"
      "scientist" { tiedemie:s+het :N42 FRONT } %<P N PL NOM
   "<or>"
      "or" { tai } %CC CC
   "<researchers>"
      "researcher" { tutkij:a+at :N12 } %<P N PL NOM +at NOM
   "<.>"
      "." { . }
```

Next we implement the front/back affiliation (20).

(20)

```
   "<*this>"
```

```
   "this" { tä:mä } %SUBJ PRON DEM SG CAPINIT
"<means>"
   "mean" { tarkoitt+aa } O-PAR %+FMAINV V PRES SG3 +aa SG
"<that>"
   "that" { , että } %CS CS
"<there>"
   "there" { NOGLOSS } LOC %F-SUBJ <Ex> ADV
"<is>"
   "be" { o+n } %+FMAINV V PRES SG3
"<a lot of>"
   "a lot of" { paljon } M-PAR %DN> DET
"<important>"
   "important" { tärke+ää } %A> A ABS SG PAR
"<research>"
   "research" { tutkimu+sta } %PCOMPL-S N SG PAR
"<,>"
   "," { , }
"<which>"
   "which" { , jo+ta } %SUBJ <Rel> PRON WH NOM SG NOM
"<is>"
   "be" { NOGLOSS } %+FAUXV V PRES SG3 PASS-PRES
"<neither>"
   "neither" { eivät } %ADVL ADV
"<seen>"
   "see" { nä+e } O-ACC %-FMAINV PL NEG-PRES
"<nor>"
   "nor" { eivätkä } %CC CC
"<read>"
   "read" { lu+e } O-PAR %-FMAINV EN-AG PL NEG-PRES
"<by>"
   "by" { NOGLOSS AG-PART } POST %ADVL PREP
"<scientists>"
   "scientist" { tiedemie+het } %<P N PL NOM
"<or>"
   "or" { tai } %CC CC
"<researchers>"
   "researcher" { tutkij+at } %<P N PL NOM
"<.>"
   "." { . }
```

In (20) we see that the word "important" is glossed as tärke+ää. That is, the two final letters 'a' have changed to 'ä', according to the conversion rules. Now each Finnish word has its final format. What still remains to be done is to change the word order. This is not a trivial task, because we have to change the passive structure into active structure. In order to be able to write reordering rules, we have to put the sentence on one line (21).

(21)

```
( DEM { tämä } %SUBJ PRON SG CAPINIT ) ( V { tarkoittaa } O-PAR
%+FMAINV PRES SG3 +aa SG ) ( CS { , että } %CS ) ( ADV { NOGLOSS }
```

```
LOC %F-SUBJ <Ex> ) ( V { on } %+FMAINV PRES SG3 ) ( DET { paljon }
M-PAR %DN> ) ( A { tärkeää } %A> ABS SG PAR ) ( N { tutkimusta }
%PCOMPL-S SG PAR ) ( { , } ) ( WH { , jota } %SUBJ <Rel> PRON NOM
SG NOM ) ( PASS-PRES { NOGLOSS } %+FAUXV V PRES SG3 ) ( ADV {
eivät } %ADVL ) ( NEG-PRES { näe } O-ACC %-FMAINV PL ) ( CC {
eivätkä } %CC ) ( NEG-PRES { lue } O-PAR %-FMAINV PL ) ( POST PREP
{ NOGLOSS AG-PART } %ADVL ) ( N { tiedemiehet } %<P PL NOM ) ( CC
{ tai } %CC ) ( N { tutkijat } %<P PL NOM +at NOM ) ( { . } )
```

Using a reordering rule, the sentence in (21) is converted into (22).

(22)

```
( DEM { tämä } %SUBJ PRON SG CAPINIT ) ( V { tarkoittaa } O-PAR
%+FMAINV PRES SG3 +aa SG ) ( CS { , että } %CS ) ( ADV { NOGLOSS }
LOC %F-SUBJ <Ex> ) ( V { on } %+FMAINV PRES SG3 ) ( DET { paljon }
M-PAR %DN> ) ( A { tärkeää } %A> ABS SG PAR ) ( N { tutkimusta }
%PCOMPL-S SG PAR ) ( { , } ) ( WH { , jota } %SUBJ <Rel> PRON NOM
SG NOM ) ( PASS-PRES { NOGLOSS } %+FAUXV V PRES SG3 ) :( N {
tiedemiehet } %<P PL NOM ) ( CC { tai } %CC ) ( N { tutkijat } %<P
PL NOM ) :( ADV { eivät } %ADVL ) ( NEG-PRES { näe } O-ACC %-
FMAINV PL ) ( CC { eivätkä } %CC ) ( NEG-PRES { lue } O-PAR %-
FMAINV PL ) ( { . } )
```

When we remove unnecessary tags, we get the final translation (23).

(23)
Tämä tarkoittaa, että on paljon tärkeää tutkimusta, jota tiedemiehet ja tutkijat eivät näe eivätkä lue.

6 The use of gerund instead of infinitive

Another type of nightmare for machine translation is that English uses gerund also in cases where infinitive would be more appropriate. Such gerund form can be verb, noun, or adjective, and it is very difficult to disambiguate such forms. An example of this problem is in (24).

(24)

```
I believe the scientific community needs to start seriously
tackling this issue.
```

When we translate this, we get the following incorrect translation (25).

(25)
Minä uskon tieteellisen yhteisön tarvitsee aloittaa vakavasti käsiksi käyminen tätä kysymystä.

Again, if the source text is Swahili, we get a correct translation. Let us see this step by step. In Swahili the sentence is as in (26).

(26)
`Nasadiki kwamba jamii ya kisayansi inahitaji kuanza kupambana na swala hilo kwa makini.`

This is first translated into English (27).

(27)
`I believe that the scientific community needs to begin to fight with this problem seriously.`

When we now translate the English version to Finnish, we get a correct translation.

Minä uskon, että tieteellisen yhteisön tarvitsee alkaa taistella tämän ongelman kanssa vakavasti.

7 Discussion

It is interesting to see the performance of Google Translate in translating our example sentences. Let us first take the original sentence (28).

(28)
Original: This means a lot of important research is not seen or read by scientists and researchers.
GT: *Tämä merkitsee paljon Tärkeää tutkimusta ei nähnyt tai lukea Tutkijat ja tutkija.*

The improved form of the sentence is translated in this way (29).

(29)
Original: This means that there is a lot of important research, which is neither seen nor read by the scientists or the researchers.
GT: *Tämä tarkoittaa, että on paljon tutkimusta Tärkeää, joka ei ole nähnyt eikä lukea Tutkijat tai tutkijan.*

And finally the same text translated from Swahili to Finnish (30).

(30)
Original: Hii inamaanisha kwamba kuna kwa wingi utafiti muhimu, ambao hauonekani wala hausomwi na wanasayansi wala watafiti.
GT: *Tämä tarkoittaa, että on runsaasti tutkimusta tarpeen, jota ei ole nähty eikä usomwi tai tutkijoita.*

The second example sentence from English to Finnish is translated in this way (31).

(31)
Original: I believe the scientific community needs to start seriously tackling this issue.

GT: *Uskon tiedeyhteisö tarvitsee aloittaa vakavasti ongelman ratkaisun.*

Translation from Swahili to Finnish is in (32).

(32)
Original: Nasadiki kwamba jamii ya kisayansi inahitaji kuanza kupambana na swala hilo kwa makini.
GT: *Uskon, että tiedeyhteisö tarvitsee aloittaakseen taistelun tätä asiaa huolellisesti*.

The detailed description above shows, that although the example sentences seem simple, they may be surprisingly difficult to translate into a language, which deviates radically from the source language. It also shows that a rule-based translation system does not encounter such translation problems that cannot be solved.

We have also learned that translation from a third language through English may bring better results than translation directly from English. This is mostly due the fact, that when translating from a language into English, the translation is grammatically correct and it retains also such features that in current English use are usually lost. Often occurring examples include the omission of relative pronouns and of conjunctions initiating a subordinate clause. Also the use of gerund instead of infinitive is a constant problem in translation. Using English as interlingua such problems can be bypassed.