HELDA
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

**https://helda.helsinki.fi**

# Translation via interlingua : Use of source language codes

Hurskainen, Arvi

University of Helsinki, Institute for Asian and African Studies

2020

Hurskainen , A 2020 ' Translation via interlingua : Use of source language codes ' Technical reports on language technology , no. 49 , University of Helsinki, Institute for Asian and African Studies , Helsinki . < http://www.njas.helsinki.fi/salama/translation-via-interlingua.pdf >

http://hdl.handle.net/10138/317511

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.*

*Technical Reports on Language Technology*
*Report No 49, 2020*
*http://www.njas.helsinki.fi/salama*

# Translation via interlingua: Use of source language codes

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

**Abstract**

In this report I will discuss issues related to translation between two languages, using interlingua as turntable. Test languages are Swahili and Finnish, both morphologically complex and linguistically very different languages. In an earlier report (Report No. 30) I discussed the issue considering the normalised version of English as interlingua. In this report the emphasis is on the use of linguistic information of the source language in various phases of the translation process. Normally, all linguistic information is lost in the translation process. Also, when the translation from interlingua to the third language is carried out, the analyser expects that the language is clean text without linguistic tags. In this report we study the possibility of retaining the linguistic information also in the analysis of interlingua (modified English). Various tag combinations are tested.

**Key Words:** *morphology, machine translation, interlingua.*

## 1 Introduction

This paper is based on the assumption, that it is possible to construct a global machine translation system between all languages using the normalised English as interlingua. The modification of English concerns mostly such anomalies, which have been introduced to the language over the past few decades. Such anomalies include the omission of sub-clause markers, the use of gerund instead of infinitive, the omission of relative clause markers etc.

When a sentence is translated from the language such as Swahili or Finnish into English, these anomalies can be avoided, because the source languages do not contain those anomalies. All those features, which are missing in current English, are clearly encoded. Therefore, there is motivation to retain this information in the translation process. Furthermore, the analysis of the source language contains a lot of such linguistic information, which would be valuable, when the translation is extended to the third language.

Because no normal analyser of English would allow the mixture of surface words and linguistic codes, I have constructed such an analyser of English, which allows the defined linguistic codes to be attached to the end of each surface word. These coded are the codes of the source language and not the codes of English.

The translation module, which translates from language A into interlingua, produces correct English word forms and the correct word order, but it adds also some linguistic codes to the end of each word. The analyser of English accepts these codes, but the codes do not affect the analysis of the surface English in any way. Yet the codes are retained in the analysis result.

The result of English analyser contains two sets of tags, those inherited from the source language and those inserted by the English analyser. As a result, we have a rich set of linguistic information, which makes the disambiguation on English vastly simpler than what it would be without tags of the source language. The syntactic tags of the source language can be accepted as such to English, and no new syntactic mapping of English is needed.

The disambiguation of English morphology can be carried out mostly by simply comparing the inherited tags and the tags added by the English analyser. The syntactic tags of the source language are there already. After disambiguation, double tagging can be reduced into single tagging, and such inherited tags that remain in the result, will be converted into the form expected by the translation system from English to Finnish. The rest of the translation process can be carried out using the normal routines.

## 2 Process phases

Below we follow in detail how the translation process proceeds. We use the example sentence *Mtu niliyemwona jana amekuja* (The man whom I saw yesterday has come).

(1)

```
"<*mtu>"
      "mtu" N 1/2-SG HUM { the } { man , :human being , person }
CAP
"<niliyemwona>"
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-SUB-REL { who } 1-
SG3-OBJ OBJ { him , her } z [ona] { see , feel } SVO HUM-ACT
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-OBJ-REL { whom } 1-
SG3-OBJ OBJ { him , her } z [ona] { see , feel } SVO HUM-ACT
"<jana>"
      "jana" N 9-SG { yesterday } TIME
      "jana" ADV { yesterday } PREFR TIME
"<amekuja>"
      "ja" V 1-SG3-SP VFIN { he } PERF:me 1-SG2-OBJ OBJ { you } z
[ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { he } PERF:me 15-SG-OBJ OBJ { it } z
[ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { he } PERF:me 17-SG-OBJ OBJ { then } z
[ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { he } PERF:me 17-SG-OBJ OBJ { there }
z [ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { he } PERF:me INFMARK z [ja] { come }
SV MONOSLB
      "ja" V 1-SG3-SP VFIN { she } PERF:me 1-SG2-OBJ OBJ { you } z
[ja] { come } SV MONOSLB
```

```
      "ja" V 1-SG3-SP VFIN { she } PERF:me 15-SG-OBJ OBJ { it } z
[ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { she } PERF:me 17-SG-OBJ OBJ { then }
z [ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { she } PERF:me 17-SG-OBJ OBJ { there }
z [ja] { come } SV MONOSLB
      "ja" V 1-SG3-SP VFIN { she } PERF:me INFMARK z [ja] { come }
SV MONOSLB
"<.$>"
      ".$" { .$ } **CLB
```

The sentence has two verbs, *ona* and *ja*. Both have prefixes attached to the verb stem. Prefixes have glosses, and each alternative gloss (e.g. *he* and *she*) prompts a new reading. This feature could also be implemented so that in the morphological lexicon we put alternative glosses on the same line and perform the splitting operation later. We see such an operation below.

When we perform morphological disambiguation, we get a result as in (2).

(2)

```
"<*mtu>"
      "mtu" N 1/2-SG HUM { the } { man } CAP @SUBJ
"<niliyemwona>"
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-OBJ-REL { whom } 1-
SG3-OBJ OBJ { him } z [ona] { see } SVO HUM-ACT @FMAINVtr-OBJ>
"<jana>"
      "jana" ADV { yesterday } PREFR TIME @ADVL
"<amekuja>"
      "ja" V 1-SG3-SP VFIN { he } PERF:me INFMARK z [ja] { come }
SV MONOSLB @FMAINVintr
"<.$>"
      ".$" { .$ } **CLB
```

This result is correct, but it has inaccurate encoding, because the presence of the subject tag and object tag in the verb depends on the sentence structure. Therefore, we must add to the reading also such alternatives (3).

(3)

```
"<*mtu>"
      "mtu" N 1/2-SG HUM { the } { man } CAP @SUBJ
"<niliyemwona>"
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-OBJ-REL { whom } 1-
SG3-OBJ OBJ { him } z [ona] { see } SVO HUM-ACT @FMAINVtr-OBJ>
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-OBJ-REL { whom } 1-
SG3-OBJ OBJ NO-OBJ-GLOSS z [ona] { see } SVO HUM-ACT @FMAINVtr-
OBJ>
      "ona" V 1-SG1-SP VFIN NO-SP-GLOSS PAST 1-SG-OBJ-REL { whom }
1-SG3-OBJ OBJ { him } z [ona] { see } SVO HUM-ACT @FMAINVtr-OBJ>
      "ona" V 1-SG1-SP VFIN NO-SP-GLOSS PAST 1-SG-OBJ-REL { whom }
1-SG3-OBJ OBJ NO-OBJ-GLOSS z [ona] { see } SVO HUM-ACT @FMAINVtr-
OBJ>
```

```
"<jana>"
      "jana" ADV { yesterday } PREFR TIME @ADVL
"<amekuja>"
      "ja" V 1-SG3-SP VFIN { he } PERF:me INFMARK z [ja] { come }
SV MONOSLB @FMAINVintr
      "ja" V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z [ja] {
come } SV MONOSLB @FMAINVintr
"<.$>"
      ".$" { .$ } **CLB
```

In (3) there are separate lines for all alternatives, which are: (a) subject and object glosses are present, (b) subject gloss is present but object gloss is absent, (c) subject gloss is absent but object gloss is present, (d) subject gloss and object gloss are absent.
This reading is then disambiguated using context information (4).

(4)

```
 "<*mtu>"
      "mtu" N 1/2-SG HUM { the } { *man } CAP @SUBJ
"<niliyemwona>"
      "ona" V 1-SG1-SP VFIN { *i } PAST 1-SG-OBJ-REL { whom } 1-
SG3-OBJ OBJ NO-OBJ-GLOSS z [ona] { see } SVO HUM-ACT @FMAINVtr-
OBJ>
"<jana>"
      "jana" ADV { yesterday } PREFR TIME @ADVL
"<amekuja>"
      "ja" V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z [ja] {
come } SV MONOSLB @FMAINVintr
"<.$>"
      ".$" { .$ } **CLB
```

We see that for the verb *ona*, the alternative with subject gloss present and object gloss absent, was chosen.

We continue in the translation process to the phase, where we can check whether all English words have the syntactic tag. Note that Swahili includes into the verb such features as pronoun subject, pronoun object, and relative pronoun, and they do not get separate syntactic mapping of Swahili, because they are not separate words. This is demonstrated in (5).

(5)

```
( N 1/2-SG HUM { *the } { *man } CAP @SUBJ )
( V 1-SG1-SP { whom } PAST 1-SG-OBJ-REL VFIN { *i } 1-SG3-OBJ OBJ
NO-OBJ-GLOSS z { see } SVO HUM-ACT @FMAINVtr-OBJ> )
( ADV { yesterday } PREFR TIME @ADVL )
( V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z { come } SV
@FMAINVintr )
( ".$" { .$ } **CLB )
```

The gloss **i* does not have a syntactic tag, and it must be added. Also, the relative prefix, glossed as *whom*, needs its syntactic tag (6).

(6)

```
( N 1/2-SG HUM { *the } { *man } CAP @SUBJ )
( V 1-SG1-SP PRON { whom } @OBJ PAST 1-SG-OBJ-REL VFIN PRON { *i }
@SUBJ 1-SG3-OBJ OBJ NO-OBJ-GLOSS z { see } SVO HUM-ACT @FMAINVtr-
OBJ> )
( ADV { yesterday } PREFR TIME @ADVL )
( V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z { come } SV
@FMAINVintr )
( ".$" { .$ } **CLB )
```

We now proceed with translation and convert the English glosses into surface form (7).

(7)

```
( N 1/2-SG HUM { *the } { man } @SUBJ )
( V 1-SG1-SP PRON { whom } @OBJ PAST 1-SG-OBJ-REL VFIN PRON { *i }
@SUBJ  1-SG3-OBJ OBJ NO-OBJ-GLOSS z { saw } SVO HUM-ACT @FMAINVtr-
OBJ> )
( ADV { yesterday } PREFR TIME @ADVL )
( V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z { has } { come }
SV @FMAINVintr )
( ".$" { .$ } **CLB )
```

This is the phase of translation, where we gradually attach tags to the English surface forms. First, we attach POS tags, and also semantic tags, if present (8).

(8)

```
( N 1/2-SG HUM { *the } { man+N } @SUBJ )
( V 1-SG1-SP { whom+PRON } @OBJ PAST 1-SG-OBJ-REL VFIN { *i+PRON }
@SUBJ  1-SG3-OBJ OBJ NO-OBJ-GLOSS z { saw+V } SVO HUM-ACT
@FMAINVtr-OBJ> )
( ADV { yesterday+ADV } PREFR TIME @ADVL )
( V 1-SG3-SP VFIN NO-SP-GLOSS PERF:me INFMARK z { has+V } { come+V
} SV @FMAINVintr )
( ".$" { .$ } **CLB )
```

Then we attach TAM tags of verbs  and semantic tags (9).

(9)

```
( N 1/2-SG { *the } { man+N+HUM } @SUBJ )
( V 1-SG1-SP { whom+PRON } @OBJ 1-SG-OBJ-REL VFIN { *i+PRON }
@SUBJ  1-SG3-OBJ OBJ NO-OBJ-GLOSS z { saw+V+PAST } SVO HUM-ACT
@FMAINVtr-OBJ> )
( ADV { yesterday+ADV } PREFR TIME @ADVL )
( V 1-SG3-SP VFIN NO-SP-GLOSS INFMARKz { has+V+PERF:me } { come+V
} SV @FMAINVintr )
( ".$" { .$ } **CLB )
```

Finally, we add syntactic tags (10).

(10)

```
( N 1/2-SG { *the } { man+N+HUM+@SUBJ } )
( V 1-SG1-SP { whom+PRON+@OBJ } 1-SG-OBJ-REL VFIN { *i+PRON+@SUBJ
} 1-SG3-OBJ OBJ NO-OBJ-GLOSSz { saw+V+PAST+@FMAINVtr-OBJ> } SVO
HUM-ACT )
( ADV { yesterday+ADV+@ADVL } PREFR TIME )
( V 1-SG3-SP VFIN NO-SP-GLOSS INFMARKz { has+V+PERF:me } {
come+V+@FMAINVintr } SV )
( ".$" { .$ } **CLB )
```

Note that all along we have retained the POS glosses also in their original places, because they are in key role in reordering of words.

Now we convert the sentence into surface form, with inherited tags attached to the words (11).

(11)

```
The man_n_hum_-subj whom_pron_-obj I_pron_-subj saw_v_past_-
fmainvtr-obj yesterday_adv_-advl has_v_perf:me come_v_-fmainvintr.
```

Note that the tags are converted to lower case, so that they can be clearly distinguished from the new upper-case tags of the English analyser. Also, some characters are temporarily rewritten to such forms, which the analyser accepts.

Now when we analyse this text form with the English analyser, we get the form as in (12).

(12)

```
"<*the>"
      "the" DET CAP

"<man_n_hum_-subj>"
      "man_n_hum_-subj" N SG
      "man_n_hum_-subj" V INF
      "man_n_hum_-subj" V PRES

"<whom_pron_-obj>"
      "who_pron_-obj" PRON REL OBJ

"<*i_pron_-subj>"
      "i_pron_-subj" PRON CAP SG1

"<saw_v_past_-fmainvtr-obj>"
      "saw_v_past_-fmainvtr-obj" V INF
      "saw_v_past_-fmainvtr-obj" V PRES PL/PRES SG1/PRES SG2
      "see_v_past_-fmainvtr-obj" V PAST

"<yesterday_adv_-advl>"
      "yesterday_adv_-advl" N SG
      "yesterday_adv_-advl" ADV
```

```
"<has_v_perf:me>"
      "have_v_perf:me" V PRES SG3

"<come_v_-fmainvintr>"
      "come_v_-fmainvintr" V INF
      "come_v_-fmainvintr" V PRES PL/PRES SG1/PRES SG2
      "come_v_-fmainvintr" V EN

"<.$>"
      "." **CLB
```

We see that some verb readings have more than one interpretation on one line, and we must produce a separate line for each alternative. The reason for allowing such under-defined readings is that by so doing we can condense the morphological lexicon. The reading suitable for disambiguation is in (13).

(13)

```
"<*the>"
      "the" DET CAP

"<man_n_hum_-subj>"
      "man_n_hum_-subj" N SG
      "man_n_hum_-subj" V INF
      "man_n_hum_-subj" V PRES

"<whom_pron_-obj>"
      "who_pron_-obj" PRON REL OBJ

"<*i_pron_-subj>"
      "i_pron_-subj" PRON CAP SG1

"<saw_v_past_-fmainvtr-obj>"
      "saw_v_past_-fmainvtr-obj" V INF
      "saw_v_past_-fmainvtr-obj" V PRES PL
      "saw_v_past_-fmainvtr-obj" V PRES SG1
      "saw_v_past_-fmainvtr-obj" V PRES SG2
      "see_v_past_-fmainvtr-obj" V PAST

"<yesterday_adv_-advl>"
      "yesterday_adv_-advl" N SG
      "yesterday_adv_-advl" ADV

"<has_v_perf:me>"
      "have_v_perf:me" V PRES SG3

"<come_v_-fmainvintr>"
      "come_v_-fmainvintr" V INF
      "come_v_-fmainvintr" V PRES PL
      "come_v_-fmainvintr" V PRES SG1
```

```
      "come_v_-fmainvintr" V PRES SG2
      "come_v_-fmainvintr" V EN

"<.$>"
      "." **CLB
```

In order to properly disambiguate the readings above, we must detach the tags as separate units (14)

(14)

```
"<*the>"
      "the" DET CAP

"<man>"
      "man" n hum @subj N SG
      "man" n hum @subj V INF
      "man" n hum @subj V PRES

"<whom>"
      "who" pron @obj PRON REL OBJ

"<*i>"
      "i" pron @subj PRON CAP SG1

"<saw>"
      "saw" v past @fmainvtr-obj V INF
      "saw" v past @fmainvtr-obj V PRES PL
      "saw" v past @fmainvtr-obj V PRES SG1
      "saw" v past @fmainvtr-obj V PRES SG2
      "see" v past @fmainvtr-obj V PAST

"<yesterday>"
      "yesterday" adv @advl N SG
      "yesterday" adv @advl ADV

"<has>"
      "have" v perf:me V PRES SG3

"<come>"
      "come" v @fmainvintr V INF
      "come" v @fmainvintr V PRES PL
      "come" v @fmainvintr V PRES SG1
      "come" v @fmainvintr V PRES SG2
      "come" v @fmainvintr V EN

"<.$>"
      ". **CLB
```

When we write disambiguation rules on the basis of this reading, we can compare the inherited tags with new tags. The obvious solution is that if the tag in both is the same, or comparable, that reading must be the correct one. If such selection is not possible, the environment helps in making the correct choice.

On the part of syntactic tags, we can see that each word has a tag. Even the words such as **i* and *whom*, which originally had no tag, have now the correct tag, because it was added later (see (6) above).

After disambiguation, the result is as in (15).

(15)
```
"<*the>"
      "the" DET CAP
"<man>"
      "man" n hum @subj N SG
"<whom>"
      "who" pron @obj PRON REL OBJ
"<*i>"
      "i" pron @subj PRON CAP SG1
"<saw>"
      "see" v past @fmainvtr-obj V PAST
"<yesterday>"
      "yesterday" adv @advl ADV
"<has>"
      "have" v perf:me V PRES SG3
"<come>"
      "come" v @fmainvintr V EN
"<.$>"
      ". **CLB
```

We see above that the inherited POS tags match with new tags, and the analysis is correct. Now we remove redundant tags and convert the remaining tags into upper-case, as the system expects (16).

(16)
```
"<*the>"
      "the" DET CAP
"<man>"
      "man" HUM @SUBJ N SG
"<whom>"
      "who" @OBJ PRON REL OBJ
"<*i>"
      "i" @SUBJ PRON CAP SG1
"<saw>"
      "see" @FMAINVtr-obj V PAST
"<yesterday>"
      "yesterday" @ADVL ADV
"<has>"
      "have" V PRES SG3
"<come>"
```

```
      "come" @FMAINVintr V EN
"<.$>"
```

Now Finnish glosses are added. Alternative glosses are on the same line. And as we see, some words have several glosses (17).

(17)
```
"<*the>"
      "the" DET CAP
"<man>"
      "man { mies N42 FRONT , ihminen N38 FRONT } HUM" HUM @SUBJ N
SG
"<whom>"
      "who" { joka Np13 } @OBJ PRON REL OBJ
"<*i>"
      "i { minä Np5 , NOGLOSS , itse N8 FRONT } HUM OUT" @SUBJ
PRON CAP SG1
"<saw>"
      "see { nähdä V71 FRONT TRV V-3INF-INE , nähdä V71 FRONT ,
katso O-PAR , on nähtävissä O-INE }" @FMAINVtr-obj V PAST
"<yesterday>"
      "yesterday { eilen }" @ADVL ADV
"<has>"
      "have { olla V67b S-ADE O-PAR TRV-N , olla V67b HAVE-PERF ,
omistaa V67 , NOGLOSS , TRV-N , en ole , en ollut , et ole , ei
ole O-PAR , ei ollut O-PAR , emme ole , ette ole , eivät ole :2 ,
eivät olleet :3 , saada V63 TRV , NOGLOSS }" perf:me V PRES SG3
"<come>"
      "come { tulla V67 O-LOC3 , tulla V67 O-LOC2 , olla V67b
kotoisin } MOVE" @FMAINVintr V EN "<.$>"
```

Each gloss is put on its own line (18).

(18)
```
"<*the>"
      "the" { *the } DET CAP
"<man>"
      "man" { mies N42 FRONT } HUM HUM @SUBJ N SG
      "man" { ihminen N38 FRONT } HUM HUM @SUBJ N SG
"<whom>"
      "who" { joka Np13 } @OBJ PRON REL OBJ
"<*i>"
      "i" { minä Np5 } HUM OUT @SUBJ PRON CAP SG1
      "i" { NOGLOSS } HUM OUT @SUBJ PRON CAP SG1
      "i" { itse N8 FRONT } HUM OUT @SUBJ PRON CAP SG1
      "i" { PROP-CAND } HUM OUT @SUBJ PRON CAP SG1
"<saw>"
      "see" { nähdä V71 FRONT TRV V-3INF-INE } @FMAINVtr-obj V
PAST
      "see" { nähdä V71 FRONT } @FMAINVtr-obj V PAST
```

```
      "see" { katso O-PAR } @FMAINVtr-obj V PAST
      "see" { on nähtävissä O-INE } @FMAINVtr-obj V PAST
"<yesterday>"
      "yesterday" { eilen } @ADVL ADV
"<has>"
      "have" { olla V67b S-ADE O-PAR TRV-N } perf:me V PRES SG3
      "have" { olla V67b HAVE-PERF } perf:me V PRES SG3
      "have" { omistaa V67 } perf:me V PRES SG3
      "have" { NOGLOSS } perf:me V PRES SG3
      "have" { TRV-N } perf:me V PRES SG3
      "have" { en ole } perf:me V PRES SG3
      "have" { en ollut } perf:me V PRES SG3
      "have" { et ole } perf:me V PRES SG3
      "have" { ei ole O-PAR } perf:me V PRES SG3
      "have" { ei ollut O-PAR } perf:me V PRES SG3
      "have" { emme ole } perf:me V PRES SG3
      "have" { ette ole } perf:me V PRES SG3
      "have" { eivät ole :2 } perf:me V PRES SG3
      "have" { eivät olleet :3 } perf:me V PRES SG3
      "have" { saada V63 TRV } perf:me V PRES SG3
      "have" { NOGLOSS } perf:me V PRES SG3
"<come>"
      "come" { tulla V67 O-LOC3 } MOVE @FMAINVintr V EN
      "come" { tulla V67 O-LOC2 } MOVE @FMAINVintr V EN
      "come" { olla V67b kotoisin } MOVE @FMAINVintr V EN
"<.$>"
```

This reading will be semantically disambiguated (19).

(19)

```
"<*the>"
      "the" { *the } DET CAP
"<man>"
      "man" { mies N42 FRONT } HUM HUM @SUBJ N SG
"<whom>"
      "who" { joka Np13 } @OBJ PRON REL OBJ
"<*i>"
      "i" { minä Np5 } HUM OUT @SUBJ PRON CAP SG1
"<saw>"
      "see" { nähdä V71 FRONT } TRV V-3INF-INE @FMAINVtr-obj V
PAST
"<yesterday>"
      "yesterday" { eilen } @ADVL ADV
"<has>"
      "have" { olla V67b } S-ADE O-PAR TRV-N perf:me V PRES SG3
"<come>"
      "come" { tulla V67 } O-LOC3 MOVE @FMAINVintr V EN
"<.$>"
```

Inflection tags are added to the reading (20).

(20)

```
"<*the>"
      "the" { *the } DET CAP @PAR
"<man>"
      "man" { mies N42 FRONT } HUM HUM SUBJ  N SG NOM
"<whom>"
      "whom" { joka Np13 }OBJ  PRON REL OBJ ACC
"<*i>"
      "i" { minä Np5 } HUM OUT SUBJ  PRON CAP SG1
"<saw>"
      "see" { nähdä V71 FRONT } TRV V-3INF-INE FMAINV tr-obj V
PAST
"<yesterday>"
      "yesterday" { eilen } ADVL  ADV
"<has>"
      "have" { olla V67b } S-ADE O-PAR TRV-N perf:me V PRES SG3
"<come>"
      "come" { tulla V67 } O-LOC3 MOVE FMAINV intr V EN EN-PERF
@SG
"<.$>"
```

Word stem boundaries are marked with colon (21).

(21)

```
"<*the>"
      "the" { *the } DET CAP PAR
"<man>"
      "man" { mie:s N42 FRONT } HUM HUM SUBJ  N SG NOM
"<whom>"
      "whom" { , jo:ka Np13 } OBJ  PRON REL OBJ SG ACC
"<*i>"
      "i" { min:ä Np5 } HUM OUT SUBJ  PRON CAP SG1
"<saw>"
      "see" { nä:hdä V71 FRONT } TRV V-3INF-INE FMAINV tr-obj V
PAST SG1
"<yesterday>"
      "yesterday" { eilen } ADVL  ADV
"<has>"
      "have" { o:lla V67b } S-ADE O-PAR TRV-N perf:me V PRES  SG
"<come>"
      "come" { tul:la V67 } O-LOC3 MOVE FMAINV intr V EN EN-PERF
SG
"<.$>"
```

The Finnish words are converted to surface form, using the information available on the inflection tags (22).

(22)

```
"<man>"
```

```
      "man" { mie:s } %SUBJ HUM DEF N SG NOM
"<whom>"
      "who" { , jo+nka } %OBJ <Rel> PRON WH SG ACC
"<I>"
      "i" { min:ä } %SUBJ HUM OUT PRON PERS SG NOM
"<saw>"
      "see" { nä+in } %+FMAINV TRV V-3INF-INE V PAST SG1
"<yesterday>"
      "yesterday" { eilen } %ADVL ADV { , }
"<has>"
      "have" { o+n } %+FAUXV S-ADE O-PAR TRV-N V PRES SG
"<come>"
      "come" { tul+lut } %-FMAINV O-LOC3 MOVE V EN-PERF SG
"<.>"
      "." { . }
```

The final translation is in (23).

(23)
*Mies, jonka minä näin eilen, on tullut.*

**3 Testing with complex syntactic forms**

The translation from language A via interlingua to language C is a complex process with several such points, where translation may go wrong. Below we test some sentence structures, where word order changes in translation.

(24)
*Watu wale niliowaona jana wamekuja.*

Interlingua:
```
Those_-ndem people_n_hum_-subj whom_pron_-obj I_pron_-subj
saw_v_past_-fmainvtr-obj yesterday_adv_-advl have_v_perf:me
come_v_-fmainvintr.
```

Target language:
*Nuo ihmiset, jotka minä näin eilen, ovat tulleet.*

(25)
*Watu wale wawili niliowaona jana wamekuja.*

Interlingua:
```
Those_-ndem two_num people_n_hum_-subj whom_pron_-obj I_pron_-subj
saw_v_past_-fmainvtr-obj yesterday_adv_-advl have_v_perf:me
come_v_-fmainvintr.
```

Target language:

*Nuo kaksi ihmistä, jotka minä näin eilen, ovat tulleet.*

(26)
*Watu wale wawili warefu niliowaona jana wamekuja.*

Interlingua:
```
Those_-ndem two_num long_adj_-nadj People_n_hum_-subj whom_pron_-
obj I_pron_-subj saw_v_past_-fmainvtr-obj yesterday_adv_-advl
have_v_perf:me come_v_-fmainvintr.
```

Target language:
*Nuo kaksi pitkää ihmistä, jotka minä näin eilen, ovat tulleet.*

(27)
*Vitu vyangu vile viwili virefu nilivyovinunua jana vimekuja.*

Interlingua:
```
Those_-advl my_sg1_-gcon two_num long_adj_-nadj things_n_-subj
which I_pron_-subj bought_v_past_-fmainvtr-obj yesterday_adv_-advl
have_v_perf:me come_v_-fmainvintr.
```

Target language:
*Ne kaksi pitkää esinettä, jotka ostin eilen, ovat tulleet.*

The tests above show that the rules, which were originally designed for handling word order in bi-lingual translation, work also in the translation system, where interlingua is used. This is possible, if we preserve the critical codes, such as the POS codes, which are made use of in rules for controlling word order.

## 4 Conclusion

The tests made in this report show that it is possible to make use of the linguistic information of the source language, when language A is translated via interlingua into language C. Both test languages are very complex, and at the same time very different morphologically and syntactically. The method makes it possible to reduce the morphological disambiguation of interlingua to minimum. And in syntax, the codes inherited from language A can be used as such. Overall, several heavy processes can be avoided.

Although the tests were made using a single development environment and similar approaches in various phases, the approach does not exclude other approaches. It allows any type of translation system to be integrated into the system. The only requirement is that the translation system is able to produce high quality translation between a non-English language and interlingua.