

Verb Chains in English to Finnish Machine Translation

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

Abstract

The main verb of a clause may have one or more modifying verbs, which describe the function of the main verb in more detail. Modal verbs are a group of verbs, which require at least one modifying verb in the clause (explicitly or as assumed). There is also a larger group of verbs, which may or may not have a modifying verb. Such verb chains in English can be translated with corresponding chains in Finnish. The problem is that the inflection of the modifying verb in target language cannot be predicted from source language. This report suggests a solution to the problem

Key Words: *machine translation, verb chain.*

1 Introduction

Verb chains are such constructions, where two or more verbs in a clause form an interconnected chain. The construction has the main verb and one or more modifying verbs. These chain structures can be divided into two groups. The first group, modal verbs, consists of verbs with limited inflection, and they require at least one modifying verb. The modifying verb can be explicit in the clause, but it also can be assumed (not explicit in the clause), whereby the context reveals which verb is meant. Consider example (1).

- (1)
(a) Should you go?
(b) Yes, I should.

In (a), the modal verb *should* has the modifying verb *go*, as the rule requires. In (b), the modifying verb is missing, but the context reveals that it is *go*.

The second group consists of verbs, which may have a modifying verb but can also be without it. Consider example (2).

- (2)
(a) I should ask him to go.
(b) I should ask him whether he will go.

In (a), the verb *ask* has a modifying verb *go*. In (b), the verb *ask* does not have a modifying verb. The verb *go* belongs to another clause. Various structures of chain verbs will be discussed below.

2 Basic verb chains and their translation

When the sentences in (2) are translated, the problem is not only the inflected form of the verbs. Also, the verb *ask* is translated with different Finnish verbs. An intermediate phase of the translation process is in (3).

(3)

```
(a) "<I>"
    "i" { minä Np5 } HUM OUT %SUBJ PRON PERS NOM SG1
"<should>"
    "should" { pitäisi } %+FAUXV V AUXMOD
"<ask>"
    "ask" { pyytää V54-F FRONT O-PAR V-3INF-ILL } %-FMAINV V INF
"<him>"
    "he" { hän Np9 FRONT OUT } HUM %OBJ PRON PERS SG3
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
    "go" { mennä V67 FRONT O-LOC3 V-3INF-ILL } MOVE %-FMAINV V
INF
"<.>"
    "." { . }
(b) "<I>"
    "i" { minä Np5 } HUM OUT %SUBJ PRON PERS NOM SG1
"<should>"
    "should" { pitäisi } %+FAUXV V AUXMOD
"<ask>"
    "ask" { kysyä V52 FRONT O-ABL } %-FMAINV V INF
"<him>"
    "he" { hän Np9 FRONT OUT } HUM %I-OBJ PRON PERS SG3
"<whether>"
    "whether" { KO } %CS CS
"<he>"
    "he" { hän Np9 FRONT OUT } HUM %SUBJ PRON PERS NOM SG3
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD
"<go>"
    "go" { mennä V67 FRONT O-LOC3 V-3INF-ILL } MOVE %-FMAINV V
INF
"<.>"
    "." { . }
```

This semantically disambiguated phase shows that in (a) the verb *ask* is translated with *pyytää*, but in (b) with *kysyä*. We also see that in (a), there is a code V-3INF-ILL meaning

that the modifying verb should be in third infinitive illative form. In (b), the verb *ask* does not have a corresponding code, because the clause does not have a modifying verb (the verb *go* is in another clause). In (4) we see how inflection tags are added.

```
(4)
(a) "<I>"
    "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
"<should>"
    "should" { pitäisi } %+FAUXV V AUXMOD SG1
"<ask>"
    "ask" { pyytää V54-F FRONT } %-FMAINV O-PAR V-3INF-ILL V INF
SG1
"<him>"
    "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
    "go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<.>"
    "." { . }
(b) "<I>"
    "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
"<should>"
    "should" { pitäisi } %+FAUXV V AUXMOD SG1
"<ask>"
    "ask" { kysyä V52 FRONT } %-FMAINV O-ABL V INF SG1
"<him>"
    "he" { hän Np9 FRONT } %I-OBJ OUT HUM PRON PERS SG3 @ABL
"<whether>"
    "whether" { KO } %CS CS
"<he>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM PRON PERS NOM SG3 @NOM
"<will>"
    "will" { NOGLOSS } %+FAUXV V AUXMOD SG @PRES
"<go>"
    "go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG @PRES
"<.>"
    "." { . }
```

Such inflection codes, which are not inherited from source language, are added with the prefix @. We see that the main subject should be in genitive. The object of *ask* in (a) should be in partitive, but in (b) in ablative. In (a), the modifying verb *go* should be in third infinitive illative. But in (b), the verb *go* is not in a modifying position, and in principle it should be interpreted as infinitive. However, when the preceding verb *will* is interpreted as future marker, the tag @PRES is added to indicate that it should have a present tense form. The sentences are translated in (5).

(5)

Minun pitäisi pyytää häntä menemään.

Minun pitäisi kysyä häneltä meneekö hän.

The verb may also have more than one modifier, whereby each modifier has the identical relationship to the main verb. Consider example (6).

(6)

```
"<I>"
    "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
"<should>"
    "should" { pitäisi } %+FAUXV V AUXMOD SG1
"<ask>"
    "ask" { pyytää V54-F FRONT } %-FMAINV O-PAR V-3INF-ILL V INF
SG1
"<him>"
    "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<sit_down>"
    "sit_down" { istuutua V52-F } %-FMAINV MW V INF SG1 @3INF-
ILL
"<,>"
    "," { , }
"<write>"
    "write" { kirjoittaa V53-C } %-FMAINV TRV O-ALL V INF SG1
@3INF-ILL
"<and>"
    "and" { ja } %CC CC
"<read>"
    "read" { lukea V58-D } %-FMAINV O-PAR V INF SG1 @3INF-ILL
"<.>"
    "." { . }
```

The translation is in (7).

(7)

Minun pitäisi pyytää häntä istuutumaan, kirjoittamaan ja lukemaan.

3 More verbs in the chain

Theoretically, there is no limit to how many verbs can be in the hierarchically ordered verb chain. In practice, however, the limit seems to be in in four verbs. An example of such a chain is in (8).

```
(8)
"<I>"
  "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
"<should>"
  "should" { pitäisi } %+FAUXV V AUXMOD SG1
"<ask>"
  "ask" { pyytää V54-F FRONT } %-FMAINV O-PAR V-3INF-ILL V INF
SG1
"<him>"
  "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
  "go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<clarify>"
  "clarify" { selventää V54-J FRONT } %-FMAINV O-PAR V INF SG1
@3INF-ILL
"<matter>"
  "matter" { asia N12 } %OBJ DEF N SG SG @PAR
"<.>"
  "." { . }
```

We see that the verb *ask* has a tag V-3INF-ILL. This tag triggers the rule, which adds the tag @3INF-ILL to the verb *go*. The verb *go* on its part has the tag V-3INF-ILL, which triggers the rule for adding the tag @3INF-ILL to the verb *clarify*. The verb *ask* has the inherited tag INF, and no modification is needed. With the aid of this tag combination, we get the clean translation (9).

```
(9)
Minun pitäisi pyytää häntä menemään selventämään asiaa.
```

So far we have seen that the inflected form of the modifying verb tends to be the third infinitive illative. However, this is not always the case. If we change the above example a bit, we get the structure as in (10).

```
(10)
"<I>"
  "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
"<should>"
  "should" { pitäisi } %+FAUXV V AUXMOD SG1
"<forbid>"
  "forbid" { kieltää V54-I FRONT } %-FMAINV O-PAR V-3INF-ELA V
INF SG1
"<him>"
  "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<from>"
```

```
"from" { M-LOC2 } %ADVL PREP
"<going>"
  "go" { mennä V67 FRONT } %<P-FMAINV O-LOC3 V-3INF-ILL MOVE V
ING SG @3INF-ELA
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<clarify>"
  "clarify" { selventää V54-J FRONT } %-FMAINV O-PAR V INF SG
@3INF-ILL
"<matter>"
  "matter" { asia N12 } %OBJ DEF N SG SG @PAR
"<.>"
  "." { . }
```

The verb *forbid* has the tag V-3INF-ELA, which triggers the tag @3INF-ELA on the verb *go*. The verb *go* on its part has the tag V-3INF-ILL, which adds the tag @3INF-ILL on the verb *clarify*. The clean translation is in (11).

(11)

Minun pitäisi kieltää häntä menemästä selventämään asiaa.

The modifying verb may also have the third infinitive inessive form. Consider the example in (12).

(12)

```
"<I>"
  "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @NOM
"<saw>"
  "see" { nähdä V71 FRONT } %+FMAINV TRV V-3INF-INE V PAST SG1
"<him>"
  "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @ACC
"<going_home>"
  "go_home" { mennä V67 FRONT kotiin } %-FMAINV MW V ING SG
@3INF-INE
"<.>"
  "." { . }
```

The tag V-3INF-INE in the verb *see* triggers the rule, which adds the tag @3INF-INE on the verb *go_home*. Note that the words *go* and *home* were interpreted as a multiword expression. The translation is in (13).

(13)

Minä näin hänet menemässä kotiin.

In the present phase of the development of the system, there are 126 such verbs, which require the third infinitive illative form from the modifying verb, 9 verbs requiring the third infinitive elative, and 4 verbs requiring the third infinitive inessive.

4 Alternative translations

In some cases it is not clear which translation is preferable, and how the modifying verb should be inflected. A typical case is the use of *alkaa/ruveta* as a translation for the verbs *begin* and *start*. Consider the examples in (14).

(14)

```
(a) "<He>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM CAPINIT PRON PERS NOM
SG3 @NOM
"<started>"
    "start" { aloittaa V53-C } %+FMAINV TRV V PAST SG
"<work>"
    "work" { työ N19 FRONT } %OBJ DEF N SG SG @ACC
"<.>"
    "." { . }

(b) "<He>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM CAPINIT PRON PERS NOM
SG3 @NOM
"<started>"
    "start" { alkaa V56-D } %+FMAINV V PAST SG
"<walking>"
    "walk" { kävellä V67 FRONT } %-FMAINV MOVE V ING SG @INF
"<.>"
    "." { . }

(c) "<He>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM CAPINIT PRON PERS NOM
SG3 @NOM
"<started>"
    "start" { ruveta V74-E } %+FMAINV V-3INF-ILL V PAST SG
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<walk>"
    "walk" { kävellä V67 FRONT } %-FMAINV MOVE V INF SG @3INF-
ILL
"<.>"
    "." { . }

(d) "<He>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM CAPINIT PRON PERS NOM
SG3 @NOM
"<began>"
    "begin" { alkaa V56-D } %+FMAINV TRV V PAST SG
"<working>"
    "work" { toimia V61 } %-FMAINV V-1INF-TRA V ING SG @INF
"<.>"
    "." { . }

(e) "<He>"
    "he" { hän Np9 FRONT } %SUBJ OUT HUM CAPINIT PRON PERS NOM
SG3 @NOM
"<began>"
```

```
"begin" { ruveta V74-E } %+FMAINV V-3INF-ILL V PAST SG
"<to>"
"to" { NOGLOSS } %INFMARK> INFMARK>
"<work>"
"work" { toimia V61 } %-FMAINV V-1INF-TRA V INF SG @3INF-ILL
"<.>"
"." { . }
```

The verbs *begin* and *start* function here as synonyms, and both can be translated with *alkaa* or *ruveta*, which also are synonyms. It is not important which one of the Finnish verbs should be used. What is important, however, is that when the verb is selected, its modifying verb must be inflected correctly. With the verb *alkaa* it is infinitive, but with *ruveta* it is third infinitive illative. Some people in Finland do not obey this rule and use also the modifying verb of *alkaa* in third infinitive illative. Recently the national language committee of Finland went so far that they publicly accepted also this form as correct. This nearly ignited a civil war and the issue is still far from settled. In the current translation system, such a heresy is not permitted, and the verb *alkaa* will never be accompanied with its modifying verb in third infinitive illative.

Because the verbs *alkaa* and *ruveta* are synonyms, it is very difficult to write a rule for controlling their selection. In order to demonstrate the behaviour of both verbs, I have above implemented the selection so that when the English verb is followed by the infinitive marker, the translation is *ruveta*. If it is in gerund, the translation is *alkaa*. The translation is in (15).

- (15)
- (a) Hän aloitti työn.
 - (b) Hän alkoi kävellä.
 - (c) Hän rupesi kävelemään.
 - (d) Hän alkoi toimia.
 - (e) Hän rupesi toimimaan.

5 Inflection of the main verb in verb chains

In verb chains, the main verb is inflected in time, also in cases, where a modal verb precedes it. Modal verbs often do not inflect at all. They just indicate the necessity or likelihood of the action, which the main verb describes. When the main verb inflects, it does not have effect on the inflection of its immediate modifying verb, or on the other modifying verbs in the verb chain. In (16) are examples of verb constructions, where the main verb inflects.

- (16)
- (a) "<I>"
"i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 @GEN
 - "<should>"
"should" { pitäisi } %+FAUXV V AUXMOD SG1
 - "<ask>"


```
"ask" { pyytää V54-F FRONT } %-FMAINV O-PAR V-3INF-ILL V INF
SG1
"<him>"
"he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
"to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
"go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<.>"
"." { . }
(b) "<I>"
"i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@GEN
"<should>"
"should" { olisi pitänyt :2 } %+FAUXV V AUXMOD SG1
"<have>"
"have" { NOGLOSS } %-FAUXV V INF SG1
"<asked>"
"ask" { pyytää V54-F FRONT } %-FMAINV O-PAR V-3INF-ILL V EN
SG @INF
"<him>"
"he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
"to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
"go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<.>"
"." { . }
(c) "<I>"
"i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@GEN
"<should>"
"should" { olisi pitänyt :2 } %+FAUXV V AUXMOD SG1
"<have>"
"have" { NOGLOSS } %-FAUXV V INF SG1
"<forbidden>"
"forbid" { kieltää V54-I FRONT } %-FMAINV O-PAR V-3INF-ELA V
EN SG @INF
"<him>"
"he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
"to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
"go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ELA
"<.>"
"." { . }
(d) "<I>"
```

```
"i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@GEN
"<should>"
  "should" { olisi pitänyt :2 } %+FAUXV V AUXMOD SG1
"<have>"
  "have" { NOGLOSS } %-FAUXV V INF SG1
"<commanded>"
  "command" { komentaa V54-J } %-FMAINV O-PAR V-3INF-ILL V EN
SG @INF
"<him>"
  "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
  "go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<sit>"
  "sit" { istua V52 } %-FMAINV O-INE V INF SG1 @3INF-ILL
"<.>"
  "." { . }
(e) "<I>"
  "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@GEN
"<should>"
  "should" { ei olisi pitänyt :5 } %+FAUXV V AUXMOD SG1 @NEG-
PRES
"<not>"
  "not" { NOGLOSS } %ADVL NEG-PART
"<have>"
  "have" { NOGLOSS } %-FAUXV V INF SG1
"<commanded>"
  "command" { komentaa V54-J } %-FMAINV O-PAR V-3INF-ILL V EN
SG @INF
"<him>"
  "he" { hän Np9 FRONT } %OBJ OUT HUM PRON PERS SG3 @PAR
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<go>"
  "go" { mennä V67 FRONT } %-FMAINV O-LOC3 V-3INF-ILL MOVE V
INF SG1 @3INF-ILL
"<to>"
  "to" { NOGLOSS } %INFMARK> INFMARK>
"<sit>"
  "sit" { istua V52 } %-FMAINV O-INE V INF SG1 @3INF-ILL
"<.>"
  "." { . }
(f) "<I>"
  "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@PAR
```

```
"<should>"
    "should" { ei olisi pitänyt :5 } %+FAUXV V AUXMOD SG1 @NEG-
PRES
"<not>"
    "not" { NOGLOSS } %ADVL NEG-PART
"<have>"
    "have" { NOGLOSS } %-FAUXV V INF SG1
"<been>"
    "be" { NOGLOSS } %-FAUXV O-LOC1 V EN SG @INF
"<commanded>"
    "command" { komentaa V54-J } %-FMAINV O-PAR V-3INF-ILL V EN
SG @INF
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<sit>"
    "sit" { istua V52 } %-FMAINV O-INE V INF SG1 @3INF-ILL
"<.>"
    "." { . }
(g) "<I>"
    "i" { minä Np5 } %SUBJ HUM OUT PRON PERS NOM SG1 CAPINIT
@PAR
"<should>"
    "should" { ei olisi pitänyt :5 } %+FAUXV V AUXMOD SG1 @NEG-
PRES
"<not>"
    "not" { NOGLOSS } %ADVL NEG-PART
"<have>"
    "have" { NOGLOSS } %-FAUXV V INF SG1
"<been>"
    "be" { NOGLOSS } %-FAUXV O-LOC1 V EN SG @INF
"<forbidden>"
    "forbid" { kieltää V54-I FRONT } %-FMAINV O-PAR V-3INF-ELA V
EN SG @INF
"<to>"
    "to" { NOGLOSS } %INFMARK> INFMARK>
"<sit>"
    "sit" { istua V52 } %-FMAINV O-INE V INF SG1 @3INF-ELA
"<.>"
    "." { . }
```

The examples above show that whatever the type of inflection of the main verb is, the modifying verb inflects in the same way. This applies also to cases with passive structure, as in (f) and (g). The translations are in (17).

(17)

- (a) *Minun pitäisi pyytää häntä menemään.*
- (b) *Minun olisi pitänyt pyytää häntä menemään.*
- (c) *Minun olisi pitänyt kieltää häntä menemästä.*
- (d) *Minun olisi pitänyt komentaa häntä menemään istumaan.*
- (e) *Minun ei olisi pitänyt komentaa häntä menemään istumaan.*

- (f) *Minua ei olisi pitänyt komentaa istumaan.*
(g) *Minua ei olisi pitänyt kieltää istumasta.*

6 Conclusion

Verb chains in English have such a structure, that the main verb, possibly preceded by a modal verb, inflects as normal, and the modifying verb or verbs are in infinite form or sometimes in gerund. This information is not sufficient for translating into Finnish, because in Finnish the modifying verb may be in infinitive, but also often in one of the third infinitive forms. Therefore, the inflection pattern of each verb must be encoded into the verb in the lexicon. With the help of this code system it is then possible to convert the English text into correct Finnish form.