Technical Reports in Language Technology
Report No 38, 20019
http://www.njas.helsinki.fi/salama

Semantic disambiguation

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

Abstract

In rule-based machine translation, semantic disambiguation is perhaps the most difficult phase to implement. Whereas the morphological disambiguation is largely based on tags and set names, in semantic disambiguation this is only partially possible. In this report we will discuss various possibilities for writing rules for semantic disambiguation.

Key Words: *semantic disambiguation, finite-state technology, machine translation*

1 Introduction

Rule-based machine translation needs a carefully designed module for semantic disambiguation. Most words have more than one meaning, and often only one of them suits for translation in a given context. Morphological disambiguation, which precedes semantic disambiguation, is a fairly simple operation, at least on morphological rich languages, because morphological features as well as syntactic relations help in disambiguation. On the semantic level the task is more difficult, because concrete features, which could ve used in formulating disambiguation rules, are nomally absent. In dictionaries it is normal to classify words into word classes, but semantic classification of words is usally absent in them.

If no semantic classification is available, the simplest approach is to proceed case by case. A disambiguation rule is written each time when need occurs. Often this is the only procedure available. The work is slow and the number of rules easily accumulates. Therefore, there is a need to write more general rules, such that would suit to a set of cases.

Given the difficulty of semantic disambiguation, it is understandable that many developers resort to statistical methods in this phase. Although statistical methods in semantic disambiguation are tempting, they have their own drawbacks. They are not reliable, because the solutions are based on guessing and not on knowing. In addition, by introducing statistican elements into the translation system we invite the well-known problems of statistical methods, the long-sentence problem and the rare-word problem. Therefore, it is worth trying to perform semantic disambiguation using overt disambiguation rules.

In this report, we shall discuss the challences of semantic disambiguation and figure out solutions for making the rule-based semantic disambiguation manageable.

A large part of semantic disambiguation is performed in the phase, where multi-word expressions are isolated. In this isolation, the MWE usually has only one interpretation.

However, there are cases, where the MWE has more than one interpretation. In such cases, semantic disambiguation must be performed later.

2 Semantic classes

Although dictionaries do not describe semantic classes, it is possible to introduce them into the system later. Obvious classes are such entities as humanness, *animacity, food, time, month, week, place, country, language, season, color, planet, title, vehicle*, etc. Also the content of text can be made use of. In some texts it is useful to know whether the text handles *medicine, law, compuer science* etc.

Semantic classes can be encoded in two ways. In one method, the members of a class are listed in the disambiguation rule file, such as in (1).

(1)
```
LIST MONTH = "january" "february" "march" "april" "may" "june"
"july" "august" "september" "october" "november" "december" MONTH
;
```

In another method, semantic tags are added to the conversion lexicon. In (2) tags are added to nouns.

(2)
```
actor; actor { näyttelijä N12 FRONT , toimija N12 } HUM;
woman; woman { nainen N38 } HUM FEM;
brakeman; brakeman { jarrumies N42 FRONT } HUM MALE;
kauklahti; kauklahti { *kauklahti N7-F } PLACE IN;
tampere; tampere { *tampere N48 } PLACE OUT;
```

In (2) we see that the class HUM is a super class, which includes the sub classes FEM and MALE. Also PLACE is a superclass, which includes the sub classes IN and OUT. The latter sub classes direct the system to use inner and outer locatives in connection with the word concerned. One says *Kauklahdessa* (IN) and *Tampereella* (OUT).

Also verbs have often tags, which help in semantic disambiguation (3).

(3)
```
advance; advance { edetä V72-F FRONT , jouduttaa V53-C TRV ,
edistää V53 FRONT O-PAR } MOVE;
```

The entry above (3) has three translation alternatives for the English verb *advance*. The verbs *jouduttaa* and *edistää* are transitive verbs, and the verb *edetä* is intransitive. One of the transitive verbs has the tag TRV, meaning that it is a transitive verb, which prefers the object case accusative. The second transitive verb has the tag O-PAR, meaning that it prefers the object case partitive. The problem of handling the object case is far from simple. The tags give just rough guidance on choosing the object case. The rules operate so that if no rule defines the object case differently, then the default case, marked on the gloss, is chosen.

Marking the transitive verbs gives the possibility to write powerful rules for semantic disambiguation. If the clause does not have an object, all transitive alternatives can be discarded. English does not mark transitivity in any way. Therefore, the presence or absence of the object gives grounds for disambiguation.
Below are examples of both cases (4).

(4)
```
"<He>"
      "he" { hän Np9 FRONT OUT } HUM MALE %SUBJ CAPINIT PRON PERS
NOM SG3
      "he" { hänen } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { NOGLOSS } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { itse N8 FRONT } HUM MALE %SUBJ CAPINIT PRON PERS NOM
SG3
      "he" { PROP-CAND } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
"<advanced>"  S:1561/4
      "advance" { edetä V72-F FRONT } MOVE %+FMAINV V PAST
"<to>"  S:122/2, 121/2
      "to" { M-LOC3 } %ADVL PREP
      "to" { NOGLOSS M-PAR } %ADVL PREP
      "to" { tai } %ADVL PREP
      "to" { - DASH } %ADVL PREP
      "to" { sinne } %ADVL PREP
"<new>"
      "new" { uusi N27 } %A> INDEF A ABS
      "new" { uudempi N16-H } %A> INDEF A ABS
      "new" { uusin N51 } %A> INDEF A ABS
"<phase>"
      "phase" { vaihe N48 } %<P INDEF N NOM SG
"<.>"
      "." { . }
```

The rule, which removed the transitive alternatives from the verb *advance*, is quite complicated, as we see in (5)

(5)
```
REMOVE OBJ-MARK-STRICT (NOT -1 BE OR ("have_be")) (NOT *1 OBJ
BARRIER CLB OR COMMA) (NOT 0 ("need") OR ("call") OR ("become") OR
("see") OR ("follow") OR ("be_call") OR ("do") OR ("love") OR EN
OR ("continue") OR ("show") OR ("shoot") OR ("get") OR (PROP-
CAND)) (NOT 1 ("with") OR ("more") OR (CS)) (NOT *-1 EN OR OBJ
BARRIER CLB OR COMMA LINK -1 BE) (NOT -2 BE);
```

The tag OBJ-MARK-STRICT is a set name, and it includes all proper object tags, such as TRV, O-PAR, O-ACC-G and O-ACC-N. The rule looks for object to both directions to make sure that an object does not occur within the same clause. However, there are verbs, for which the rule should not apply. The list in the rule is not exhaustive. They are just the verbs encountered to have problems, and certainly there are more of them.
The translation is in (6).

(6)
Hän eteni uuteen vaiheeseen.

Below (7) is an example, where the clause has an object.

(7)
```
"<He>"
      "he" { hän Np9 FRONT OUT } HUM MALE %SUBJ CAPINIT PRON PERS
NOM SG3
      "he" { hänen } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { NOGLOSS } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { itse N8 FRONT } HUM MALE %SUBJ CAPINIT PRON PERS NOM
SG3
      "he" { PROP-CAND } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
"<has>"  S:1565/4
      "have" { olla V67b HAVE-PERF } %+FAUXV V PRES SG3
      "have" { omistaa V67 } %+FAUXV V PRES SG3
      "have" { NOGLOSS } %+FAUXV V PRES SG3
      "have" { en ole } %+FAUXV V PRES SG3
      "have" { en ollut } %+FAUXV V PRES SG3
      "have" { et ole } %+FAUXV V PRES SG3
      "have" { emme ole } %+FAUXV V PRES SG3
      "have" { ette ole } %+FAUXV V PRES SG3
      "have" { eivät ole :2 } %+FAUXV V PRES SG3
      "have" { eivät olleet :3 } %+FAUXV V PRES SG3
"<advanced>"  S:1566/4
      "advance" { jouduttaa V53-C TRV } MOVE %-FMAINV V EN
      "advance" { edistää V53 FRONT O-PAR } MOVE %-FMAINV V EN
"<process>"
      "process" { prosessi N5 } %OBJ DEF N SG
"<.>"
      "." { . }
```

The intransitive alternative for the verb *advance* has been removed, and both transitive alternatives are left. The verbs *jouduttaa* and *edistää* have similarities in meaning, but they are not full synonyms. The verb *jouduttaa* means speeding up, while *edistää* refers to promoting. In this sentence, the first alternative is appropriate, and no further rule is needed. The translation is in (8)

(8)
Hän on jouduttanut prosessin.

The translation is not quite correct, because the object should be in partitive. The gloss jouduttaa has the tag TRV, and it should be changed to O-PAR. If we do this, we get the correct translation (9)

(9)
Hän on jouduttanut prosessia.

The example shows how difficult it is to assign the correct default object case for a verb. In order to test the second transitive alternative, we must have a sentence, where the translation fits. Consider the example in (10).

(10)
```
"<Obviously>"
      "obviously" { todennäköisesti STI } %ADVL CAPINIT ADV
      "obviously" { todennäköisen } %ADVL CAPINIT ADV
      "obviously" { ilmeisesti STI } %ADVL CAPINIT ADV
      "obviously" { ilmeisen } %ADVL CAPINIT ADV
      "obviously" { tietenkin } %ADVL CAPINIT ADV
      "obviously" { PROP-CAND } %ADVL CAPINIT ADV
"<,>"  S:190/4
      "," { NOGLOSS }
"<he>"
      "he" { hän Np9 FRONT OUT } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { hänen } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { NOGLOSS } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { itse N8 FRONT } HUM MALE %SUBJ PRON PERS NOM SG3
"<has>"  S:1566/4
      "have" { olla V67b HAVE-PERF } %+FAUXV V PRES SG3
      "have" { omistaa V67 } %+FAUXV V PRES SG3
      "have" { NOGLOSS } %+FAUXV V PRES SG3
      "have" { en ole } %+FAUXV V PRES SG3
      "have" { en ollut } %+FAUXV V PRES SG3
      "have" { et ole } %+FAUXV V PRES SG3
      "have" { emme ole } %+FAUXV V PRES SG3
      "have" { ette ole } %+FAUXV V PRES SG3
      "have" { eivät ole :2 } %+FAUXV V PRES SG3
      "have" { eivät olleet :3 } %+FAUXV V PRES SG3
"<advanced>"  S:1497/4
      "advance" { edistää V53 FRONT O-PAR } MOVE %-FMAINV V EN
"<science>"
      "science" { tiede N48-F FRONT } SCI %OBJ DEF N SG
"<.>"
      "." { . }
```

Here the rule looks for the tag SCI (science), and if it finds it, the rule applies. The rule is in (11)

(11)
```
SELECT ("advance" V53) (*1 (SCI) BARRIER CLB);
```

The translation is in (12).

(12)
Todennäköisesti hän on edistänyt tiedettä.

3 Punctuation and auxiliary verbs

Languages have different styles to punctuate sentences. Especially the use of comma varies significantly between languages. English has commas in places where Finnish has not. English omits commas in places, where Finnish would add them. Handling this variation requires semantic disambiguation rules.

A typical case, where English uses a comma, is after the adverb in the beginning of the sentence. Finnish does not use a comma in those cases. One rule for handling such cases is in (13).

(13)
```
SELECT ("," NOGLOSS) (-1 ("for example") OR ("in_addition") OR
("if_necessary") OR ("however") OR ("therefore") OR ("instead") OR
("generally") OR ("in_general") OR ("in_total") OR ("obviously"));
```

The rule states that if a comma is immediately after the listed words, it should be removed. Although such adverbs are normally in the beginning of the sentence, the rule does not require it. It applies also elsewhere in the sentence.

This rule is not general enough for handling all similar cases. Perhaps a better rule would be in (14).

(14)
```
SELECT ("," NOGLOSS) (NOT *-1 V OR NUM BARRIER CLB);
```

The rule states that the comma should be removed, if on the left before clause boundary there is not a verb or number.

Another problem with punctuation is that sometimes English omits the comma in places where Finnish uses it. Typical cases are subordinate and relative clauses starting with the conjunction that (15). Even worse, English often omits the conjunction altogether.

(15)
(a) *He said that he will come tomorrow.*
(b) *He said he will come tomorrow.*
(c) *We cannot control speculation that is outside of the house.*

In (a), the conjunction *that* initiates a subordinate clause. But as we can see in (b), it can be omitted. In both cases there is no comma. In (c), the relative pronoun *that* initiates a relative clause, and also there is no comma.

When translating these sentences into Finnish, a comma must be added. In (b), also the conjunction *että* must be added.

It is important to note that the operations required in the examples (15) cannot be handled with disambiguation rules, because disambiguation involves selection or removal, not adding something, which is not present in the analysed text.

The system translates the above sentences as in (16).

(16)
(a) *Hän sanoi, että hän tulee huomenna.*
(b) *Hän sanoi, että hän tulee huomenna.*
(c) *Me emme voi kontrolloida spekulointia, joka on ulkopuolella talon*.

Let us see in more detail the translation process in each of the three cases.

(16a)

```
"<He>"
      "he" { hän Np9 FRONT OUT } HUM MALE %SUBJ CAPINIT PRON PERS
NOM SG3
      "he" { hänen } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { NOGLOSS } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
      "he" { itse N8 FRONT } HUM MALE %SUBJ CAPINIT PRON PERS NOM
SG3
      "he" { PROP-CAND } HUM MALE %SUBJ CAPINIT PRON PERS NOM SG3
"<said>"
      "say" { sanoa V52 TRV O-ALL SVOO } V-INF-PART HUM %+FMAINV V
PAST
      "say" { todeta V74-F TRV } V-INF-PART HUM %+FMAINV V PAST
"<that>"
      "that" { , että } %CS CS
      "that" { että :2 } %CS CS
      "that" { , jotta } %CS CS
      "that" { , joka Np13 } %CS CS
      "that" { NOGLOSS M-GEN } %CS CS
      "that" { , jolloin } %CS CS
"<he>"
      "he" { hän Np9 FRONT OUT } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { hänen } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { NOGLOSS } HUM MALE %SUBJ PRON PERS NOM SG3
      "he" { itse N8 FRONT } HUM MALE %SUBJ PRON PERS NOM SG3
"<will>"  S:1382/4
      "will" { NOGLOSS } %+FAUXV V AUXMOD
"<come>"
      "come" { tulla V67 O-LOC3 } MOVE %-FMAINV V INF
      "come" { tulla V67 O-LOC2 } MOVE %-FMAINV V INF
      "come" { olla V67b kotoisin } MOVE %-FMAINV V INF
"<tomorrow>"
      "tomorrow" { huomenna } %ADVL ADV
      "tomorrow" { huominen N38 } %ADVL ADV
"<.>"
      "." { . }
```

We see that the default gloss of the conjunction *that* is *että*, preceded by comma. Therefore, we do not need any rule for adding the comma.

For (16b), we need a rule for adding the comma and conjunction, because the English sentence has omitted both.

(16b)

```
"<He>"
      "he" { hän Np9 FRONT } %SUBJ OUT HUM MALE CAPINIT PRON PERS
NOM SG3
"<said>"  S:128/0
      "say" { sanoa V52 } %+FMAINV TRV O-ALL SVOO V-INF-PART HUM V
PAST SG  S:128 { , että }
"<he>"
      "he" { hän Np9 FRONT } %SUBJ OUT HUM MALE PRON PERS NOM SG3
"<will>"
      "will" { NOGLOSS } %+FAUXV V AUXMOD SG
"<come>"
      "come" { tulla V67 } %-FMAINV O-LOC3 MOVE V INF SG
"<tomorrow>"
      "tomorrow" { huomenna } %ADVL ADV
"<.>"
      "." { . }
```

We see that the rule has added a new gloss *{ , että },* which will be part of translation. The rule for doing this is in (17).

(17)

```
MAP ({ , että }) TARGET V (0 COMM) (NOT 0 (hallita) OR EN) (*1
VFIN BARRIER CLB OR COMMA) (*-1 SUBJ OR (IMP) BARRIER CLB) (NOT 1
COMMA OR (CC) OR EN OR NEG OR OBJ OR (%I-OBJ) OR A OR (GEN) OR
PREP OR ("there")) (NOT *2 OBJ BARRIER CLBV) (NOT -2 (<Rel>) OR
("if")) (NOT -1 COMMA OR REL);
```

Adding something, for which there is no precise criteria in text, is prone to errors. In this case we assume that the verb expresses communication in some way. Therefore, we have created a set of such verbs and given the set name COMM. The set is in (18).

(18)

```
LIST COMM = "say" "speak" "tell" "suggest" "demand" "ask" "urge"
"shout" "whisper" "think" "dream" "ensure" "guess" "rule" "find"
"argue" "assume";
```

In (19) we have an example, where the word *that* initiates a relative clause. In translation, a comma must be added in front of it. This is done already in assigning the gloss to the relative pronoun. This was already done in the phase of morphological disambiguation, and no semantic disambiguation is needed.

(19)

```
"<We>"
      "we" { me Np6 FRONT OUT } HUM %SUBJ CAPINIT PRON PERS NOM
PL1
      "we" { meidän } HUM %SUBJ CAPINIT PRON PERS NOM PL1
      "we" { NOGLOSS } HUM %SUBJ CAPINIT PRON PERS NOM PL1
      "we" { itse N8 FRONT } HUM %SUBJ CAPINIT PRON PERS NOM PL1
```

```
      "we" { PROP-CAND } HUM %SUBJ CAPINIT PRON PERS NOM PL1
"<can_not>"  S:462/4
      "can_not" { emme voi } %+FAUXV MW V AUXMOD
"<control>"
      "control" { kontrolloida V62 O-PAR } %-FMAINV V INF
      "control" { valvoa V52 O-PAR } %-FMAINV V INF
"<speculation>"
      "speculation" { spekulointi N5-J } %OBJ INDEF N SG
      "speculation" { keinottelu N2 } %OBJ INDEF N SG
      "speculation" { spekulaatio N3 } %OBJ INDEF N SG
      "speculation" { arvailu N2 } %OBJ INDEF N SG
      "speculation" { pohdiskelu N2 } %OBJ INDEF N SG
"<that>"
      "that" { , joka Np13 } %SUBJ <Rel> PRON
"<is>"  S:1562/4
      "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" { 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" { 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
"<outside_of>"
      "outside_of" { ulkopuolella } PREP M-GEN MW
      "outside_of" { ulkopuolelta } PREP M-GEN MW
      "outside_of" { ulkopuolelle } PREP M-GEN MW
"<house>"
      "house" { talo N1 } %<P DEF N NOM SG
      "house" { parlamentti N5-C } %<P DEF N NOM SG
"<.>"
      "." { . }
```

Me emme voi kontrolloida spekulointia, joka on ulkopuolella talon.

We saw in (15b) that the word *that* in the role of a conjunction can be omitted. The omission often happens, when the word *that* is in the role of a relative pronoun. An example is in (20).

(20)

```
"<person>"  S:148/0
      "person" { henkilö N2 FRONT } %SUBJ HUM DEF N NOM SG SG
S:148 { , joka Np13 } %OBJ <REL> PAR SG
"<treated>"
      "treat" { hoitaa V53-F } %-FMAINV O-PAR V EN SG
"<in>"
      "in" { NOGLOSS M-ADE } %ADVL PREP
"<this>"
```

```
      "this" { tämä Np1 FRONT } %DN> DET DEM SG
"<way>"
      "way" { tapa N9-E } %SUBJ OUT N NOM SG SG
"<must>"
      "must" { täytyy } %+FAUXV S-GEN %AUX V AUXMOD SG
"<be>"
      "be" { NOGLOSS } %-FMAINV O-LOC1 V INF SG
"<rescued>"
      "rescue" { pelastaa V53 } %-FMAINV TRV V EN SG
```

The gloss of the (missing) relative pronoun is added to the word, to which the pronoun refers. It is glossed as object, and because the main verb has a tag O-PAR, it is translated with partitive case (PAR).

The rule, which adds the reletive pronoun, is in (21).

(21)
```
MAP ({ , joka Np13 } %OBJ <REL> PAR SG) TARGET SUBJ + SG (1 (O-
PAR) + EN) (*2 VFIN BARRIER CLB) (NOT 1 (WH) OR REL OR ("call") +
EN OR (perustuen)) (NOT 0 (PERS));
```

Translation is in (22).

(22)
Henkilö, jota hoidettiin tällä tavalla, täytyy pelastaa.

We see that the relative clause is embedded in the main clause. Therefore, a comma is added also after the embedded clause (23).

(23)
```
"<person>"  S:849/0
      "person" { henkilö N2 FRONT } %SUBJ HUM DEF N NOM SG SG
S:849 @NOM
"<which>"  S:262/0
      "which" { , joka Np13 } %OBJ <REL> PAR SG S:262 @PAR
"<treated>"  S:1017/0
      "treat" { hoitaa V53-F } %-FMAINV O-PAR V EN SG  S:1017
@PASS-PAST
"<in>"
      "in" { NOGLOSS M-ADE } %ADVL PREP
"<this>"
      "this" { tämä Np1 FRONT } %DN> DET DEM SG
"<way>"  S:351/0, 136/0
      "way" { tapa N9-E } %SUBJ OUT N NOM SG SG  S:136 { , } S:351
@ADE
"<must>"
      "must" { täytyy } %+FAUXV S-GEN %AUX V AUXMOD SG
"<be>"
      "be" { NOGLOSS } %-FMAINV O-LOC1 V INF SG
```

```
"<rescued>"  S:1025/0
      "rescue" { pelastaa V53 } %-FMAINV TRV V EN SG  S:1025 @INF
"<.>"
      "." { . }
```

The rule for adding the comma is in (24).

(24)
```
ADD ({ , }) TARGET N OR (MW) OR ADV (1 VFIN) (*-1 REL BARRIER
CLB);
```

Note that the system incorrectly added the relative pronoun which, although the referent is human. However, in this translation system it does not matter, because Finnish makes no difference between humans and non-humans in relative pronouns. The solution just simplifies rule writing.

If the verb has transitive and intransitive meanings, it is often not possible to know what the correct meaning is. Consider the example in (25).

(25)
```
"<message>"
      "message" { sanoma N10 } %SUBJ DEF N NOM SG
      "message" { viesti N5 FRONT } %SUBJ DEF N NOM SG
"<hidden>"
      "hide" { kätkeä V58 FRONT TRV O-ILL } %-FMAINV V EN
      "hide" { kätkeytyä V52-I FRONT O-ILL } %-FMAINV V EN
      "hide" { piilottaa V53-C TRV } %-FMAINV V EN
      "hide" { salata V73 TRV } %-FMAINV V EN
      "hide" { peitellä V67-C FRONT TRV } %-FMAINV V EN
      "hide" { peittää V53-C FRONT TRV } %-FMAINV V EN
      "hide" { piilotella V67-C TRV } %-FMAINV V EN
      "hide" { piilottautua V52-F } %-FMAINV V EN
"<in>"  S:758/4
      "in" { NOGLOSS M-ILL } %ADVL PREP
"<jargon>"
      "jargon" { ammattikieli N26 FRONT } %<P N NOM SG
      "jargon" { slangi N5 } %<P N NOM SG
      "jargon" { ammattislangi N5 } %<P N NOM SG
      "jargon" { jargon N5 } %<P N NOM SG
      "jargon" { kapulakieli N26 } %<P N NOM SG
"<is>"  S:1566/4
      "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" { 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" { 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
```

```
"<difficult>"
      "difficult" { vaikea N15 } %PCOMPL-S INDEF A ABS
      "difficult" { vaikeampi N16-H } %PCOMPL-S INDEF A ABS
      "difficult" { vaikein N51 } %PCOMPL-S INDEF A ABS
      "difficult" { hankala N10 } %PCOMPL-S INDEF A ABS
"<to>"
      "to" { NOGLOSS } %INFMARK> INFMARK>
"<understand>"
      "understand" { ymmärtää V54-K FRONT TRV } %-FMAINV V INF
"<.>"
      "." { . }
```

No semantic disambiguation rule was applied to the verb *hide*, because the meaning is ambiguous. As a result, we get the default translation (26).

(26)
Sanoma, joka kätkettiin ammattikieleen, on vaikea ymmärtää.

Perhaps a better translation would be as in (27).

(27)
Sanoma, joka on kätketty ammattikieleen, on vaikea ymmärtää.

4 Semantic disambiguation of individual words

A large part of semantic disambiguation rules consists of specific rules with very limited possibilities to generalisation. In (28) is an extract of the rule file.

(28)

```
SELECT ("column" kolonna) (*1 (MIL) BARRIER SNTB);
SELECT ("column" kolonna) (*-1 (MIL) BARRIER SNTB);
SELECT ("rest" loput) (0 (DEF)) (1  VFIN);
SELECT ("heavily" rankasti) (*-1  ("snow") OR ("rain") BARRIER
CLB);
SELECT ("fall" sataa) (*-1  ("snow") BARRIER CLB);
SELECT ("place" sijaita) (NOT *1 OBJ BARRIER CLB);
SELECT ("player" soitin) (*1 ("cassette") OR ("record") BARRIER
SNTB);
SELECT ("player" soitin) (*-1 ("cassette") OR ("record") BARRIER
SNTB);
SELECT ("exercise" harjoittelu) (*1 (DIS) OR (MED) BARRIER SNTB);
SELECT ("exercise" harjoitella) (*1 (DIS) OR (MED) BARRIER SNTB);
SELECT ("exercising" harjoitteleminen) (*1 (DIS) OR (MED) BARRIER
SNTB);
SELECT ("heavy" rankka) (1 ("rain") OR ("rainfall") OR
("snowfall"));
SELECT ("carry" olla) (*-1 ("shop") OR ("store") BARRIER SNTB);
SELECT ("depression" masennus) (*1 (DIS) OR (MED) OR ("exercise")
OR ("exercising") BARRIER SNTB);
```

```
SELECT ("depression" masennus) (*-1 (DIS) OR (MED) OR ("exercise")
OR ("exercising") BARRIER SNTB);
```

Usually every word has more than one meaning. Therefore, the semantic disambiguation requires cross-checking the whole sentence. Consider the example in (29).

(29)

```
"<Records>"
      "record" { ennätys N39 FRONT } %SUBJ CAPINIT N PL NOM
      "record" { tietue N48 } %SUBJ CAPINIT N PL NOM
      "record" { muistiinpano N1 } %SUBJ CAPINIT N PL NOM
      "record" { merkintä N9-J FRONT } %SUBJ CAPINIT N PL NOM
      "record" { ennätys-- COMP } %SUBJ CAPINIT N PL NOM
      "record" { äänilevy N1 :2 FRONT } %SUBJ CAPINIT N PL NOM
      "record" { PROP-CAND } %SUBJ CAPINIT N PL NOM
"<are>"
      "be" { olla V67b BE TRV-N V-4INF-TRA } O-LOC1 %+FAUXV V PRES
PL
      "be" { olla V67b V-3INF-ILL } O-LOC1 %+FAUXV V PRES PL
      "be" { olla V67b V-3INF-INE } O-LOC1 %+FAUXV V PRES PL
      "be" { olla V67b BE O-PAR } O-LOC1 %+FAUXV V PRES PL
      "be" { eivät ole O-PAR V-4INF-TRA :2 } O-LOC1 %+FAUXV V PRES
PL
      "be" { eivät olleet O-PAR V-4INF-TRA :3 } O-LOC1 %+FAUXV V
PRES PL
      "be" { emme :6 } O-LOC1 %+FAUXV V PRES PL
      "be" { emme ole V-3INF-INE } O-LOC1 %+FAUXV V PRES PL
      "be" { emme olleet V-3INF-INE } O-LOC1 %+FAUXV V PRES PL
      "be" { ei ollut V-4INF-TRA } O-LOC1 %+FAUXV V PRES PL
      "be" { ei ole O-PAR V-4INF-TRA } O-LOC1 %+FAUXV V PRES PL
      "be" { NOGLOSS } O-LOC1 %+FAUXV V PRES PL
      "be" { joka Np13 } O-LOC1 %+FAUXV V PRES PL
      "be" { jotka Np14 } O-LOC1 %+FAUXV V PRES PL
      "be" { tulla V67 V-3INF-ILL } O-LOC1 %+FAUXV V PRES PL
"<played>"
      "play" { pelata V73 O-PAR } %-FMAINV V EN
      "play" { leikkiä V61-A FRONT O-PAR } %-FMAINV V EN
      "play" { näytellä V67-C FRONT O-PAR } %-FMAINV V EN
      "play" { soittaa V53-C O-PAR } %-FMAINV V EN
"<with>"
      "with" { kanssa M-GEN POST } %ADVL PREP
      "with" { mukana M-GEN POST } %ADVL PREP
      "with" { NOGLOSS M-ADE POST } %ADVL PREP
      "with" { NOGLOSS M-INE POST } %ADVL PREP
      "with" { NOGLOSS M-ILL POST } %ADVL PREP
      "with" { , jolla on O-ACC-N } %ADVL PREP
      "with" { , joilla on O-ACC-N :2 } %ADVL PREP
      "with" { NOINFL } %ADVL PREP
      "with" { NOGLOSS M-ELA } %ADVL PREP
      "with" { NOGLOSS M-ESS } %ADVL PREP
      "with" { NOGLOSS M-KOM } %ADVL PREP
```

```
"<player>"
      "player" { pelaaja N10 } HUM %<P INDEF N NOM SG
      "player" { soitin N33-C } HUM %<P INDEF N NOM SG
      "player" { toimija N12 } HUM %<P INDEF N NOM SG
```

The word *record* is disambiguated with the following rule (30).

(30)
```
SELECT ("record" :2) (*1 ("buy") OR ("purchase") OR ("sell") OR
("collect") OR ("collecting") OR ("collection") OR ("emotional")
OR ("on_sale") OR ("player") BARRIER CLB );
```

We need a rule also for the word *play* (31).

(31)
```
SELECT ("play" soittaa) (*1 ("radio") OR ("player") BARRIER SNTB);
```

The rule for the word *player* is in (32).

(32)
```
SELECT ("player" soitin) (*-1 ("cassette") OR ("record") BARRIER
SNTB);
```

The disambiguated sentence is in (33).

(33)
```
"<Records>"  S:1100/4
      "record" { äänilevy N1 :2 FRONT } %SUBJ CAPINIT N PL NOM
"<are>"  S:390/4
      "be" { NOGLOSS } O-LOC1 %+FAUXV V PRES PL
"<played>"  S:1060/4
      "play" { soittaa V53-C O-PAR } %-FMAINV V EN
"<with>"
      "with" { NOGLOSS M-ADE POST } %ADVL PREP
"<player>"  S:1512/4
      "player" { soitin N33-C } %<P INDEF N NOM SG
```

The translation is in (34).

(34)
Äänilevyt soitetaan soittimella.

Let us consider a sentence, which is slightly different (35).

(35)
```
"<Football>"
      "football" { jalkapallo N1 } SPO %A> CAPINIT N NOM SG
      "football" { jalkapalloilu N2 } SPO %A> CAPINIT N NOM SG
```

```
      "football" { jalkapallon-- COMP } SPO %A> CAPINIT N NOM SG
      "football" { PROP-CAND } SPO %A> CAPINIT N NOM SG
"<players>"
      "player" { pelaaja N10 HUM SPO } %SUBJ N PL NOM
      "player" { soitin N33-C } %SUBJ N PL NOM
      "player" { toimija N12 } %SUBJ N PL NOM
"<played>"
      "play" { pelata V73 O-PAR SPO } %+FMAINV V PAST
      "play" { leikkiä V61-A FRONT O-PAR } %+FMAINV V PAST
      "play" { näytellä V67-C FRONT O-PAR } %+FMAINV V PAST
      "play" { soittaa V53-C O-PAR } %+FMAINV V PAST
"<record>"
      "record" { ennätys N39 FRONT SPO } %A> DEF N NOM SG
      "record" { tietue N48 } %A> DEF N NOM SG
      "record" { muistiinpano N1 } %A> DEF N NOM SG
      "record" { merkintä N9-J FRONT } %A> DEF N NOM SG
      "record" { ennätys-- COMP SPO } %A> DEF N NOM SG
      "record" { äänilevy N1 :2 FRONT } %A> DEF N NOM SG
"<scores>"
      "score" { pistemäärä N10 FRONT } %OBJ DEF N PL
      "score" { joukko N1-A } %OBJ DEF N PL
"<yesterday>"
      "yesterday" { eilen } %ADVL ADV
```

When we run this through the disambiguation system, we get the result as in (36).

(36)

```
"<Football>"  S:1446/4
      "football" { jalkapallon-- COMP } SPO %A> CAPINIT N NOM SG
"<players>"
      "player" { pelaaja N10 HUM SPO } %SUBJ N PL NOM
      "player" { soitin N33-C } %SUBJ N PL NOM
      "player" { toimija N12 } %SUBJ N PL NOM
"<played>"
      "play" { pelata V73 O-PAR SPO } %+FMAINV V PAST
      "play" { leikkiä V61-A FRONT O-PAR } %+FMAINV V PAST
      "play" { näytellä V67-C FRONT O-PAR } %+FMAINV V PAST
      "play" { soittaa V53-C O-PAR } %+FMAINV V PAST
"<record>"  S:1446/4
      "record" { ennätys-- COMP SPO } %A> DEF N NOM SG
"<scores>"
      "score" { pistemäärä N10 FRONT } %OBJ DEF N PL
      "score" { joukko N1-A } %OBJ DEF N PL
"<yesterday>"
      "yesterday" { eilen } %ADVL ADV
```

A single rule has applied to two words. This rule joins the word to the following word. The rule is in (37).

(37)
```
SELECT (COMP) (1 N) (NOT 0 (CMP) OR GEN) (NOT 1 (MW) OR (CAP) OR
("time") OR (N-ING)) (NOT 2 N);
```

The translation is in (38).

(38)
Jalkapallonpelaajat pelasivat ennätyspistemääriä eilen.

5 Conclusion

Semantic disambiguation belongs to the group of problems in rule-based machine translation, where full success cannot be achieved. Problems costitute a wide spectrum. When clearly defined semantic clusters can be formed, disambiguation can be ccurate. In cases, where semantic groups cannot be isolated, the task is much more difficult.

In sum, three main strategies for optimizing semantic disambiguation can be identified. (a) The use of default translation reduces the number of rules. The first translation candidate is chosen, if no rule defines otherwise. (b) The use of set names reduces the number of needed rules. (c) For the rest of cases, rules need be written. Also here, sets should be used whenever feasible.