Technical Reports on Language Technology
Report No 79, 2022
http://www.njas.helsinki.fi/salama

Translating polysemous expressions[1]

Arvi Hurskainen
Department of Languages
FIN-00014 University of Helsinki, Finland
arvi.hurskainen@helsinki.fi
DOI: 10.13140/RG.2.2.34173.84961

Abstract

Polysemy is one of the most intricate features in machine translation. A word may have more than one meaning, and these meanings range from near-synonyms to entirely different concepts. Various machine translation systems have varying success rates in translating such structures. Google Translate (GT), the leading translation system, has constantly improved its performance in translating such structures. Its success is obviously based on model translations found in the Web. On the other hand, if such model translations are not found, GT fails badly.

GT and other leading machine translation systems use currently neural methods, and it is practically impossible to trace the translation process phase by phase. Therefore, the work with neural approaches is counter-intuitive, because we cannot fix the translation mistakes in the correct phase of the translation process.

Here I will discuss the translation of polysemic expressions in the context of rule-based machine translation. The translation system is modular, composed of several clearly ordered modules. This makes it possible to correct the mistakes on the optimal point of the translation chain.

Examples are from English, Finnish, and German.

Key Words: *polysemous words, machine translation, rule-based machine translation.*

1 Introduction

The phenomenon that a word has more than one meaning is called polysemy. Because this phenomenon is very common in languages, especially in English, correct translation becomes difficult. In the rule-based MT system, the selection of the correct meaning of each polysemous word is made using the information available in the sentence. Sometimes such crucial information is found beyond the sentence boundary, and also such information can be used. The information consists of anything found in the analysis result. Often the translation requires the use of semantic tags that must be included into the analysis lexicon.

In the rule-based translation system, the text is first analysed in great detail. This means that each grammatically correct interpretation is given to each word-form, without considering the context. The linguistic information is expressed in the form of various labels, commonly called tags. The tags are of two types. One type includes the information on part-of-speech and inflection. Another tag type consists of labels for marking semantic features, such as humanness, animacy, plants etc. The tags provide the information, on the basis of which disambiguation can be carried out.

The neural approach and rule-based approach have very little in common. The neural approach relies on surface strings and their combinations without being able to make use of grammatical information. The rule-based approach has a linguistically motivated interpretation for each word-form. This is possible, because the lexicon and the morphophonemic rules are man-made and tested by humans.

I will demonstrate the translation of polysemy with examples from Finnish and English. I also compare the translation with the translation of Google Translate.

2 Polysemic expressions in Finnish

First, we take a Finnish sentence, where the word *voi* is in three different roles with separate meanings (1)

(1)

```
"<*voi>"
      "voida" V CAP VMOD PRES SG3
      "voida" V CAP VMOD IMP SG2
      "voida" V CAP VMOD NEG-PRES
      "voi" N CAP SG ACC-N
      "voi" EXCLAM CAP
"<,>"
      "," COMMA
"<voi>"
      "voida" V VMOD PRES SG3
      "voida" V VMOD IMP SG2
      "voida" V VMOD NEG-PRES
      "voi" N SG ACC-N
      "voi" EXCLAM
"<voi>"
      "voida" V VMOD PRES SG3
      "voida" V VMOD IMP SG2
      "voida" V VMOD NEG-PRES
      "voi" N SG ACC-N
      "voi" EXCLAM
"<sulaa>"
      "sulaa" V INF
      "sulaa" V PRES SG3
      "sula" A SG PAR
"<.>"
      "." **CLB
```

We see that each of the three occurrences of the word *voi* has precisely the same interpretation. It can be an exclamation, verb, or noun.

The disambiguation system used the following Constraint Grammar rules :[2]

For *voi* as an exclamation (2)

(2)
```
SELECT ("voi" EXCLAM) (0 (CAP));
REMOVE (NEG-PRES) (NOT *-1 (NEG) BARRIER CLB);
```

For *voi* as a noun (3)

(3)
```
SELECT (N) (0 V) (NOT *1 (3INF-ILL) BARRIER CLBV) (NOT -1 ("olla")
OR INF OR PRON OR (PROPN)) (NOT *-1 (NEG) OR (VMOD) BARRIER CLB)
(NOT 0 ("sata") OR ("uhka") OR ("asu") OR ("oppi") OR ("lähetti")
OR ("panta") OR ("seura") OR ("tuli") OR ("kirja") OR ("teuras")
OR ("istua") OR ("tie") OR (2INF)) (NOT 1 INF);
REMOVE (IMP SG2) (1C (COND) OR (PRES) OR (IMP)) (NOT -1 PRON);
REMOVE (NEG-PRES) (NOT *-1 (NEG) BARRIER CLB);
```

For *voi* as a verb (4)

(4)
```
REMOVE (IMP SG2) (-1 INF OR N OR ("oleva")) (NOT -1 (Heur) OR
(ALL));
REMOVE (NEG-PRES) (NOT *-1 (NEG) BARRIER CLB);
SELECT ("voida") (1 V) (NOT 0 (3INF)) (NOT 1 (VMOD));
```

For *sulaa* as infinitive (5)

(5)
```
SELECT (INF) (-1 ("voida"));
```

After disambiguation, syntactic mapping and conversion into English, the sentence is in (6).

(6)
```
"<*voi>"
      "voi" { Oh } EXCLAM CAP @ADVL
"<,>"
      "," COMMA
"<voi>"
      "voi { butter }" N SG ACC-N @SUBJ
```

[2] For explanation of how to read CG rules see the Technical Report No. 59, http://www.njas.helsinki.fi/salama/disambiguation-strategy-of-english-text.pdf

```
"<voi>"
      "voida { can }" V VMOD PRES SG3 @+FMAINV
"<sulaa>"
      "sulaa { melt }" V INF @-FMAINV
"<.>"
      "." **CLB
```

We test how GT translates the sentence (7).

(7)
Voi, voi voi sulaa.
Oh, butter can melt.

The translation is correct.

Now we take an advertisement of the Finnish Centre Party in the election campaign in 2022 (8).

(8)
Ei ihan missä sattuu vaan siellä missä sattuu.

This sentence is quite cryptic, and it can only be understood in the context of the changing administrative structure in Finland. For a number of years, the Finnish government has tried to solve the problem of making health care and social services in Finland more efficient and cost-saving. The solution is to establish a separate administrative level between the central and local governments. The Centre Party tries to distribute services widely to the countryside, even to places with sparse population.

There are two identical word pairs *missä sattuu,* but their meaning is entirely different. The first set *missä sattuu* is an idiom and means 'wherever'. The word *sattuu* is here a necessary part of an idiom, and it cannot be translated alone. The words *siellä missä sattuu* can be translated literally as 'there where it hurts'. The message is: Don't distribute services recklessly without planning; distribute them there where patients are.

We first analyse the sentence (9)

(9)

```
"<*ei>"
      "ei" NEG CAP SG
"<ihan>"
      "ihan" ADV
"<missä>"
      "mikä PRON REL SG INE
      "mikä QUEST SG INE
      "mikä QUEST PL INE
"<sattuu>"
      "sattua" V PRES SG3
"<vaan>"
      "vaan" CONJ
"<siellä>"
```

```
      "siellä ADV
"<missä>"
      "mikä PRON REL SG INE
      "mikä QUEST SG INE
      "mikä QUEST PL INE
"<sattuu>"
      "sattua" V PRES SG3
"<.>"
      "." **CLB
```

Then we disambiguate it (10). The only word with ambiguity is *missä*.

(10)
```
"<*ei>"
      "ei" NEG CAP SG
"<ihan>"
      "ihan" ADV
"<missä>"
      "mikä" PRON REL SG INE
"<sattuu>"
      "sattua" V PRES SG3
"<vaan>"
      "vaan" CONJ
"<siellä>"
      "siellä" ADV
"<missä>"
      "mikä" PRON REL SG INE
"<sattuu>"
      "sattua" V PRES SG3
"<.>"
      "." **CLB
```

The rule for disambiguation is in (11).

(11)
```
SELECT ("mikä" REL INE) (-1 ADV);
```

Before we convert the glosses into English, we isolate the MWE *missä sattuu.* The isolation result is in (12).

(12)
```
"<*ei>"
      "ei" NEG CAP SG
"<ihan>"
      "ihan" ADV
"<missä_sattuu>"
      "mikä_sattua" { wherever } ADV
"<vaan>"
      "vaan" CONJ
"<siellä>"
```

```
      "siellä" ADV
"<missä>"
      "mikä" PRON REL SG INE
"<sattuu>"
      "sattua" V PRES SG3
"<.>"
      "." **CLB
```

Then we add English glosses (13).

(13)
```
"<*ei>"
      "ei { not }" NEG CAP SG @ADVL
"<ihan>"
      "ihan { quite }" ADV @ADVL
"<missä_sattuu>"
      "mikä_sattua" { wherever } ADV @ADVL
"<vaan>"
      "vaan { but }" CONJ @CS
"<siellä>"
      "siellä { there }" ADV @ADVL
"<missä>"
      "mikä { where }" PRON REL SG INE @ADVL
"<sattuu>"
      "sattua { hurt }" V PRES SG3 @+FMAINV
"<.>"
      "." **CLB
```

Finally, we convert the glosses into surface form, making use of tags (14).

(14)
```
"<*ei>"
      "ei { Not }" NEG CAP SG @ADVL
"<ihan>"
      "ihan { quite }" ADV @ADVL
"<missä_sattuu>"
      "mikä_sattua" { wherever } ADV @ADVL
"<vaan>"
      "vaan { but }" CONJ @CS
"<siellä>"
      "siellä { there }" ADV @ADVL
"<missä>"
      "mikä { where }" PRON REL SG INE @ADVL
"<sattuu>"
      "sattua { it hurts }" V PRES SG3 @+FMAINV
"<.>"
      "." **CLB
```

For understanding fully what the sentence means one should know the political situation in Finland in 2022 and the campaign for electing members to the board of the middle level administration.

GT translates the expression:

Ei ihan missä sattuu vaan siellä missä sattuu.
Not exactly where it hurts but where it hurts.

As we see, GT misses the point in the expression.

3 Polysemic expressions in English

Next, we take an example that has often been used for demonstrating that MT cannot translate everything. The example is a slightly modified piece of text from Bible (Mathew 26:41 and Mark 14:38) (15).

(15)
```
"<The>"
      "the" %DN> CAPINIT DET
"<spirit>"
      "spirit" %SUBJ N SG NOM DEF
"<is>"
      "be" %+FMAINV V PRES SG3
"<strong>"
      "strong" %PCOMPL-S A ABS INDEF
"<but>"
      "but" %CC CC
"<the>"
      "the" %DN> DET
"<flesh>"
      "flesh" %SUBJ N SG NOM DEF
"<is>"
      "be" %+FMAINV V PRES SG3
"<weak>"
      "weak" %PCOMPL-S A ABS INDEF
"<.>"
      "."
```

In the sentence above, the words *spirit* and *flesh* are ambiguous, as we will see when we add Finnish glosses (16).

(16)
```
"<The>"
      "the" %DN> CAPINIT DET
"<spirit>"
      "spirit" { viina , henki } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
```

```
"<strong>"
      "strong" { vahva , voimakas } %PCOMPL-S A ABS INDEF
"<but>"
      "but" { mutta } %CC CC
"<the>"
      "the" %DN> DET
"<flesh>"
      "flesh" { pihvi , liha } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<weak>"
      "weak" { heikko } %PCOMPL-S A ABS INDEF
"<.>"
      "."
```

In order to disambiguate, we need to cascade ambiguous readings (17).

(17)

```
"<The>"
      "the" %DN> CAPINIT DET
"<spirit>"
      "spirit" { viina } %SUBJ N SG NOM DEF
      "spirit" { henki } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<strong>"
      "strong" { vahva } %PCOMPL-S A ABS INDEF
"<but>"
      "but" { mutta } %CC CC
"<the>"
      "the" %DN> DET
"<flesh>"
      "flesh" { pihvi } %SUBJ N SG NOM DEF
      "flesh" { liha } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<weak>"
      "weak" { heikko } %PCOMPL-S A ABS INDEF
      "weak" { huono } %PCOMPL-S A ABS INDEF
"<.>"
      "."
```

We see that it is not possible to disambiguate the sentence without further information. The string *spirit is strong* can be translated as ‘viina on vahvaa’ or ‘henki on vahva’. Similarly, the string *flesh is weak* can be translated as ‘pihvi on huonoa’ or ‘liha on heikkoa’. We must add some context (18).

(18)

```
"<Watch>"
      "watch" { valvoa } %+FMAINV CAPINIT V IMP
```

```
"<and>"
      "and" { ja } %CC CC
"<pray>"
      "pray" { rukoilla } %+FMAINV V IMP
"<so>"
      "so" { niin } %ADVL ADV
"<that>"
      "that" { että } %CS CS
"<you>"
      "you" { sinä } %SUBJ PRON PERS NOM
      "you" { te } %SUBJ PRON PERS NOM
"<will>"
      "will" %+FAUXV V AUXMOD
"<not>"
      "not" { ei } %ADVL NEG-PART
"<fall>"
      "fall" { joutua } %-FMAINV V INF
"<into>"
      "into" %ADVL PREP
"<temptation>"
      "temptation" { kiusaus } %<P N SG NOM
"<.>"
      "."
"<The>"
      "the" %DN> CAPINIT DET
"<spirit>"
      "spirit" { viina } %SUBJ N SG NOM DEF
      "spirit" { henki } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<strong>"
      "strong" { vahva } %PCOMPL-S A ABS INDEF
"<but>"
      "but" { mutta } %CC CC
"<the>"
      "the" %DN> DET
"<flesh>"
      "flesh" { pihvi } %SUBJ N SG NOM DEF
      "flesh" { liha } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<weak>"
      "weak" { heikko } %PCOMPL-S A ABS INDEF
      "weak" { huono } %PCOMPL-S A ABS INDEF
"<.>"
      "."
```

Now we can perform disambiguation (19).

(19)

```
"<Watch>"
```

```
      "watch" { valvoa } %+FMAINV CAPINIT V IMP
"<and>"
      "and" { ja } %CC CC
"<pray>"
      "pray" { rukoilla } %+FMAINV V IMP
"<so>"
      "so" { niin } %ADVL ADV
"<that>"
      "that" { että } %CS CS
"<you>"
      "you" { sinä } %SUBJ PRON PERS NOM
"<will>"
      "will" %+FAUXV V AUXMOD
"<not>"
      "not" { ei } %ADVL NEG-PART
"<fall>"
      "fall" { joutua } %-FMAINV V INF
"<into>"
      "into" %ADVL PREP
"<temptation>"
      "temptation" { kiusaus } %<P N SG NOM
"<.>"
      "."
"<The>"
      "the" %DN> CAPINIT DET
"<spirit>"
      "spirit" { henki } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<strong>"
      "strong" { vahva } %PCOMPL-S A ABS INDEF
"<but>"
      "but" { mutta } %CC CC
"<the>"
      "the" %DN> DET
"<flesh>"
      "flesh" { liha } %SUBJ N SG NOM DEF
"<is>"
      "be" { olla } %+FMAINV V PRES SG3
"<weak>"
      "weak" { heikko } %PCOMPL-S A ABS INDEF
"<.>"
      "."
```

The CG rules that performed the disambiguation are in (20).

(20)

```
SELECT ("spirit" henki) (*-1 ("pray"));
SELECT ("flesh" liha) (*-1 ("pray"));
SELECT ("weak" heikko) (*-1 ("flesh") BARRIER CLB LINK *-1
("pray"));
```

We see that each of the rules refers to the string "pray*"* on the left. No barrier is required, and the scanning can continue until the string "pray" is found.

The third rule first finds the string "flesh", without crossing clause boundary, and then continues across the sentence boundary and finds the string "pray".

GT translates the sentence:

The spirit is strong but the flesh is weak.
Henki on vahva, mutta liha heikko.

GT picks the translation directly from the Bible translation, adds a comma, and omits the last verb.

There is also a commonly known phrase, where the ambiguous words right and left are used. I first heard the phrase applied to President Bush (21).

(21)
President Bush has two sides in his head, left and right. On the left there is nothing right, and on the right there is nothing left.

In the current some-time, one could easily be accused of racism or derogation if the subject of the expression would belong to any group, which may feel to be oppressed on any ground. For example, President Obama as a subject would be a some mine. Also, any female as subject would arouse some rage. The only safe subject is a white male. And to be fully safe, I put myself as the subject. The revised sentence is analysed in (22).

(22)

```
"<*hurskainen>"
      "hurskainen" PROPN CAP Heur
"<has>"
      "have" AUXV PRES SG3
"<two>"
      "two" NUM CARD
"<sides>"
      "side" V vt PRES SG3
      "side" N PL
"<in>"
      "in" PREP
      "in" ADV
"<his>"
      "he" PRON MALE SG3 GEN
"<head>"
      "head" V vt PRES SG1
      "head" V vt PRES SG2/PL2
      "head" V vt PRES PL1
      "head" V vt PRES PL3
      "head" V vt INF
```

```
      "head" V vt IMP
      "head" N SG
"<,>"
      "," COMMA
"<left>"
      "leave" V vt PAST
      "leave" V vt EN
      "left" A
      "left" ADV
"<and>"
      "and" CONJ CC
"<right>"
      "right" N SG
      "right" A
      "right" ADV
"<.>"
      "." **CLB
"<*on>"
      "*on" PREP CAP
      "*on" ADV CAP
"<the>"
      "the" DET DEF
"<left>"
      "leave" V vt PAST
      "leave" V vt EN
      "left" A
      "left" ADV
"<there>"
      "there" ADV
"<is>"
      "be" V PRES SG3
"<nothing>"
      "nothing" ADV
"<right>"
      "right" N SG
      "right" A
      "right" ADV
"<,>"
      "," COMMA
"<and>"
      "and" CONJ CC
"<on>"
      "on" PREP
      "on" ADV
"<the>"
      "the" DET DEF
"<right>"
      "right" N SG
      "right" A
      "right" ADV
```

```
"<there>"
      "there" ADV
"<is>"
      "be" V PRES SG3
"<nothing>"
      "nothing" ADV
"<left>"
      "leave" V vt PAST
      "leave" V vt EN
      "left" A
      "left" ADV
"<.>"
      "." **CLB
```

The word *right* has noun, adjective and adverb interpretations. The word *left* has verb, adjective and adverb interpretations. The disambiguated sentence is in (23).

(23)

```
"<*hurskainen>"
      "hurskainen" PROPN CAP Heur
"<has>"
      "have" AUXV PRES SG3
"<two>"
      "two" NUM CARD
"<sides>"
      "side" N PL
"<in>"
      "in" PREP
"<his>"
      "he" PRON MALE SG3 GEN
"<head>"
      "head" N SG
"<,>"
      "," COMMA
"<left>"
      "left" ADV
"<and>"
      "and" CONJ CC
"<right>"
      "right" A
"<.>"
      "." **CLB
"<*on>"
      "*on" PREP CAP
"<the>"
      "the" DET DEF
"<left>"
      "left" ADV
"<there>"
      "there" ADV
```

```
"<is>"
      "be" V PRES SG3
"<nothing>"
      "nothing" ADV
"<right>"
      "right" ADV
"<,>"
      "," COMMA
"<and>"
      "and" CONJ CC
"<on>"
      "on" PREP
"<the>"
      "the" DET DEF
"<right>"
      "right" ADV
"<there>"
      "there" ADV
"<is>"
      "be" V PRES SG3
"<nothing>"
      "nothing" ADV
"<left>"
      "left" ADV
"<.>"
      "." **CLB
```

The CG rules needed for disambiguation are in (24).

(24)

```
SELECT (PREP) (1 A OR PRON OR NUM) (2 N);
SELECT (N) (-1C (GEN)) (0 V) (NOT *-2 (VMOD) BARRIER CLB);
SELECT A (0 N) (*-1 V BARRIER CLB LINK NOT 0 TRV) (1 CLB);
SELECT (ADV) (NOT 1 N) (NOT 2 N) (0 A);
```

We add Finnish glosses to words (25).

(25)

```
"<*hurskainen>"
      "hurskainen" { Hurskainen } PROPN CAP Heur
"<has>"
      "have" { olla } AUXV PRES SG3
"<two>"
      "two" { kaksi } NUM CARD
"<sides>"
      "side" { puoli }N PL
"<in>"
      "in" PREP
"<his>"
      "he" { hän , NOGLOSS } PRON MALE SG3 GEN
```

```
"<head>"
      "head" { pää }N SG
"<,>"
      "," COMMA
"<left>"
      "left" { vasen } ADV
"<and>"
      "and" { ja } CONJ CC
"<right>"
      "right" { oikea } A
"<.>"
      "." **CLB
"<*on>"
      "*on" PREP CAP
"<the>"
      "the" DET DEF
"<left>"
      "left" { vasen } ADV
"<there>"
      "there" ADV
"<is>"
      "be" { olla } V PRES SG3
"<nothing>"
      "nothing" { ei mitään } ADV
"<right>"
      "right" {oikein } ADV
"<,>"
      "," COMMA
"<and>"
      "and" { ja } CONJ CC
"<on>"
      "on" PREP
"<the>"
      "the" DET DEF
"<right>"
      "right" { oikea } ADV
"<there>"
      "there" ADV
"<is>"
      "be" { olla } V PRES SG3
"<nothing>"
      "nothing" { ei mitään } ADV
"<left>"
      "left" { jäljellä }ADV
"<.>"
      "." **CLB
```

Some semantic disambiguation is needed. The possessive pronoun *hänen* has two semantic alternatives. Because it refers to the subject on the left, we choose the NOGLOSS alternative.

The translation is in (26).
Hurskaisella on kaksi puolta päässään, vasen ja oikea. Vasemmalla ei ole mikään oikein ja oikealla ei ole mitään jäljellä.

GT translates the sentence:

Hurskainen has two sides in his head, left and right. On the left there is nothing right and on the right there is nothing left.
Hurskaisen päässä on kaksi puolta, vasen ja oikea. Vasemmalla ei ole mitään oikeaa ja oikealla ei ole mitään jäljellä.

The translation is quite good, although it is not literally correct.

4 Polysemic expressions in German

In German, polysemy is alleviated by the convention to write nouns with capital-initial letters. This distinguishes them from verbs with the same polyphony. In fact, polyphony is more common in German than polysemy. Here I will take two examples, which are often used to demonstrate the similarity of words. One of them is in (27).

(27)
```
"<*der>"
      "der"  D CAPefArt SG
"<*leutnant>"
      "leutnant"  N CAP SG HUM
"<von>"
      "von"  PREP
      "von"  PrepName
"<*leuten>"
      "leute"  N CAP PL HUM ACC1
      "leute"  N CAP PL HUM ACC2
      "leuten"  PROPN
"<befahl>"
      "befehlen"  V PAST SG
"<seinen>"
      "sein"  PRON PL DAT
"<*leuten>"
      "leute"  N CAP PL HUM DAT
      "leuten"  PROPN
"<nicht>"
      "nicht"  ADV NEG
"<eher>"
      "früh"  ADV COMP
"<zu>"
      "zu"  InfMark
      "zu"  PREP
"<läuten>"
      "läuten"  N SG NOM
      "läuten"  N SG ACC1
```

```
      "läuten"  V INF
"<bis>"
      "bis"  ADV
"<der>"
      "der"  DefArt SG
"<*leutnant>"
      "leutnant"  N CAP SG HUM
"<von>"
      "von"  PREP
      "von"  PrepName
"<*leuten>"
      "leute"  N CAP PL HUM DAT
      "leuten"  PROPN
"<seinen>"
      "sein"  PRON PL DAT
"<*leuten>"
      "leute"  N CAP PL HUM DAT
      "leuten"  PROPN
"<das>"
      "das"  DefArt SG
"<*läuten>"
      "läuten"  N CAP SG NOM
      "läuten"  N CAP SG ACC
      "läuten"  V CAP INF
"<befahl>"
      "befehlen"  V PAST SG
"<.>"
      "."  **CLB
```

We see that despite the phonological similarity, the words *Leuten* (propname), *Leuten* (noun), *läuten* (verb), and *Läuten* (noun) have at least partly different spelling. However, it is a tricky sentence to disambiguate. The result is in (28).

(28)

```
"<*der>"
      "der" CAP DefArt" SG
"<*leutnant>"
      "leutnant" N CAP SG HUM
"<von>"
      "von" PREP
      "von" PREP PrepName
"<*leuten>"
      "leuten" PROPN
"<befahl>"
      "befehlen" V PAST SG
"<seinen>"
      "sein" PRON PL DAT
"<*leuten>"
      "leute" N CAP PL HUM DAT
      "leuten" PROPN
"<nicht>"
```

```
      "nicht" ADV NEG
"<eher>"
      "fruh" ADV COMP
"<zu>"
      "zu" InfMark
"<läuten>"
      "läuten" V INF
"<bis>"
      "bis" ADV
"<der>"
      "der" DefArt" SG
"<*leutnant>"
      "leutnant" N CAP SG HUM
"<von>"
      "von" PREP
      "von" PREP PrepName
"<*leuten>"
      "leuten" PROPN
"<seinen>"
      "sein" PRON PL DAT
"<*leuten>"
      "leute" N CAP PL HUM DAT
      "leuten" PROPN
"<das>"
      "das" DefArt" SG
"<*läuten>"
      "läuten" V CAP INF
"<befahl>"
      "befehlen" V PAST SG
"<.>"
      "." **CLB
```

The CG rules needed for disambiguation are in (29).

(29)

```
SELECT ("leuten" PROPN) (-1 ("von")) (-2 N);
SELECT (PREP) (1 N) (NOT -1 (GEN));
REMOVE (N) (0 INF) (*-1 VFIN BARRIER CLB) (NOT -1 (DefArt));
```

Note that the verb *befehlen* requires a dative form from the first object. The second object is in accusative form. Some forms, such as *von*, and *Leuten* (in the sense ‘people’), need no rules, because they are in the default position (first in the cohort).

Next, we add English glosses (30).

(30)

```
"<*der>"
      "der"  CAP DefArt SG
"<*leutnant>"
      "leutnant { lieutenant }"  N CAP SG HUM
```

```
"<von>"
      "von"  PREP
"<*leuten>"
      "leuten { Leuten }"  PROPN
"<befahl>"
      "befehlen { order }"  V PAST SG
"<seinen>"
      "sein { his }"  PRON DAT
"<*leuten>"
      "leute { people }"  N CAP PL HUM DAT
"<nicht>"
      "nicht { not }"  ADV NEG
"<eher>"
      "fruh { earlier }"  ADV COMP
"<zu>"
      "zu"  InfMark
"<läuten>"
      "läuten { ring }"  V INF
"<bis>"
      "bis { until }"  ADV
"<der>"
      "der"  DefArt SG
"<*leutnant>"
      "leutnant { lieutenant }"  N CAP SG HUM
"<von>"
      "von"  PREP
"<*leuten>"
      "leuten { Leuten }"  PROPN
"<seinen>"
      "sein { his }"  PRON DAT
"<*leuten>"
      "leute { people }"  N CAP PL HUM DAT
"<das>"
      "das"  DefArt SG
"<*läuten>"
      "läuten { ringing }"  N CAP SG ACC
"<befahl>"
      "befehlen { order }"  V PAST SG
"<.>"
      "."  **CLB
```

After converting the glosses to surface form, we get the translation (31).

(31) *The lieutenant from Leuten ordered his people not earlier to ring until the lieutenant from Leuten ordered to his people the ringing.*

GT translates:
Der Leutnant von Leuten befahl seinen Leuten nicht eher zu läuten bis der Leutnant von Leuten seinen Leuten das Läuten befahl.

The people's lieutenant ordered his people not to ring until the people's lieutenant ordered his people to ring.

The string The *Leutnant von Leuten* was mistranslated twice. Otherwise, the translation is correct. The solution of GT cannot be understood on any linguistic grounds. If it is a genitive construction, the original text should be *Der Leutnant der Leute*, but this is not the case. If the article *der* were omitted, the expression could also be translated as *Lieutenant von Leuten*, whereby the first part is a title and the last part the name of the person. However, the article means that *Leuten* is a place name.

We make another test and modify the original sentence by removing the definite article before *Leutnant*. The disambiguated result of this modified sentence is in (32).

(32)

```
"<*leutnant>"
      "leutnant" N CAP SG HUM
"<von>"
      "von" PREP PrepName
"<*leuten>"
      "leuten" PROPN
"<befahl>"
      "befehlen" V PAST SG
"<seinen>"
      "sein" PRON PL DAT
"<*leuten>"
      "leute" N CAP PL HUM DAT
      "leuten" PROPN
"<nicht>"
      "nicht" ADV NEG
"<eher>"
      "fruh" ADV COMP
"<zu>"
      "zu" InfMark
"<läuten>"
      "läuten" V INF
"<bis>"
      "bis" ADV
"<*leutnant>"
      "leutnant" N CAP SG HUM
"<von>"
      "von" PREP PrepName
"<*leuten>"
      "leuten" PROPN
"<seinen>"
      "sein" PRON PL DAT
"<*leuten>"
      "leute" N CAP PL HUM DAT
      "leuten" PROPN
"<das>"
      "das" DefArt" SG
```

```
"<*läuten>"
      "läuten" V CAP INF
"<befahl>"
      "befehlen" V PAST SG
"<.>"
      "." **CLB
```

We needed to add the following CG rule:

```
SELECT ("von" PrepName) (1 (N)) (NOT -2 (DefArt));
```

When we add English glosses, we get the result as in (33).

(33)

```
"<*leutnant>"
      "leutnant { lieutenant }" N CAP SG HUM
"<von>"
      "von { von }" PREP PrepName
"<*leuten>"
      "leuten { Leuten }" PROPN
"<befahl>"
      "befehlen { order }" V PAST SG
"<seinen>"
      "sein { his }" PRON PL DAT
"<*leuten>"
      "leute { people }" N CAP PL HUM DAT
"<nicht>"
      "nicht { not }" ADV NEG
"<eher>"
      "fruh { earlier }" ADV COMP
"<zu>"
      "zu" InfMark
"<läuten>"
      "läuten { ring }" V INF
"<bis>"
      "bis { until }" ADV
"<*leutnant>"
      "leutnant { lieutenant }" N CAP SG HUM
"<von>"
      "von { von }" PREP PrepName
"<*leuten>"
      "leuten { Leuten }" PROPN
"<seinen>"
      "sein { his }" PRON PL DAT
"<*leuten>"
      "leute { people }" N CAP PL HUM DAT
"<das>"
      "das" DefArt" SG
"<*läuten>"
      "läuten { ring }" V CAP INF
```

```
"<befahl>"
      "befehlen { order }" V PAST SG
"<.>"
      "." **CLB
```

After converting the glosses to surface form and after reordering words, we have the translation as in (34).

(34)
Lieutenant von Leuten ordered his people not earlier to ring until Lieutenat von Leuten ordered to his people the ringing.

GT translates this modified sentence :

Leutnant von Leuten befahl seinen Leuten nicht eher zu läuten bis Leutnant von Leuten seinen Leuten das Läuten befahl.
Lieutenant of People did not order his people to ring until Lieutenant of People ordered his people to ring.

GT makes two cardinal mistakes. First, it does not recognise that *von Leuten* is here a proper name of a person. Second, it moves negation to the left to the wrong place, which makes the sentence fluent, but changes the meaning radically.

This is typical to Google Translate. One has to ask: is it better to produce a fluent translation although the meaning changes than to produce a less fluent translation, which retains the correct meaning?

In (34) is another German expression with mostly polyphonic words. The problems are primarily syntactic, because the same words are repeated without much clue for disambiguation.

(34)

```
"<*wenn>"
      "wenn" CONJ CAP
"<hinter>"
      "hinter" PREP
"<*fliegen>"
      "fliege" N CAP PL
      "fliegen" V CAP INF
      "fliegen" V CAP PRES PL
"<*fliegen>"
      "fliege" N CAP PL
      "fliegen" V CAP INF
      "fliegen" V CAP PRES PL
"<fliegen>"
      "fliege" N PL
      "fliegen" V INF
      "fliegen" V PRES PL
```

```
"<fliegen>"
      "fliege" N PL
      "fliegen" V INF
      "fliegen" V PRES PL
"<*fliegen>"
      "fliege" N CAP PL
      "fliegen" V CAP INF
      "fliegen" V CAP PRES PL
"<*fliegen>"
      "fliege" N CAP PL
      "fliegen" V CAP INF
      "fliegen" V CAP PRES PL
"<nach>"
      "nach" VerbPref POST
"<.>"
      "." **CLB
```

There is the word *fliegen* six times as a single string. Four of them start with a capital letter, and two with a lower case letter. All of them have the same interpretation, except that the nouns have also the tag CAP.
The verb *fliegen* appears two times after each other. In fact, the latter should have the basic form *nachfliegen*, but in this sentence the prefix *nach* is detached and moved to the end.'
The disambiguated form is in (35).

(35)

```
"<*wenn>"
      "wenn"  CONJ CAP
"<hinter>"
      "hinter"  PREP
"<*fliegen>"
      "fliege"  N CAP PL
"<*fliegen>"
      "fliege"  N CAP PL
"<fliegen>"
      "fliegen"  V PRES PL
"<fliegen>"
      "fliegen"  V PRES PL
"<*fliegen>"
      "fliege"  N CAP PL
"<*fliegen>"
      "fliege"  N CAP PL
"<nach>"
      "nach"  VerbPref POST
"<.>"
      "."  **CLB
```

The sentence was disambiguated using the rules as in (36).

(36)
```
REMOVE (V) (-1C PREP);
SELECT (N CAP) (1 VFIN) (*-1 CONJ BARRIER CLB);
SELECT VFIN (NOT -1 ("zu")) (NOT 0 (CAP));
REMOVE (V) (1C POST) (NOT 1 PERS);
```

We add glosses to disambiguated words (37).

(37)
```
"<*wenn>"
      "wenn { when }" CONJ CAP @ADVL
"<hinter>"
      "hinter { behind }" PREP @ADVL
"<*fliegen>"
      "fliege { fly }" N CAP PL @<P
"<*fliegen>"
      "fliege { fly }" N CAP PL @SUBJ
"<fliegen>"
      "fliegen { fly }" V PRES PL @+FMAINV
"<fliegen>"
      "fliegen { fly }" V PRES PL @+FMAINV
"<*fliegen>"
      "fliege { fly }" N CAP PL @SUBJ
"<*fliegen>"
      "fliege { fly }" N CAP PL @SUBJ
"<nach>"
      "nach { after }" VerbPref" POST @ ADVL
"<.>"
      "." **CLB
```

All six words have precisely the same gloss *fly*. However, when we use the tags provided, we get the translation as in (38).

(38)
When behind flies flies fly flies fly after flies.

GT translates the sentence:

Wenn hinter Fliegen Fliegen fliegen fliegen Fliegen Fliegen nach.
When behind flies flies fly flies fly flies after.

GT did not recognise that the postposition should be moved before the noun in translation.

5 Conclusion

In the report I have tested the disambiguation and translation of such expressions, which have polysemic word-forms. Most of such expressions can be disambiguated on the basis

of the information available in the analysed form of the sentence. The sample has also one such case, where extra-sentential information is needed, if heuristics is avoided.

The result of the rule-based approach is compared with the translation of Google Translate. The latter has improved the translation result over the years, and in some cases the translation is correct. However, it still makes incomprehensible errors especially in places of ambiguity.