



Figure 4:10-11 (cont.)

Lund University, Dept. of Linguistics
Working Papers 33 (1988), 51-64

Some Problems for Machine Translation

Barbara Cairns

Abstract

An analysis of the translation of a Swedish magazine article text into English. Special emphasis is on the type of semantic problems which could arise in machine translation, as envisaged by a human translator. Some syntactic problems of word order are also discussed.

INTRODUCTION

The purpose of this paper is to analyse a text from the point of view of a human translator and to present some of the linguistic obstacles a machine might encounter in carrying out the same translation.

The Swedish text chosen is a magazine article, since this represents a relatively neutral use of language. Since the magazine – *Tempus* – aims at a general readership, the language is not too specialised or tailored for a specific group of readers. A more literary text would require a totally different kind of analysis and it is now generally acknowledged that the translation of literary texts is beyond the scope of any machine in existence today, as opposed to the naive and unrealistic enthusiasm of earlier machine translation work:

And now we must come to a question which has long lain in wait for us. Will the machine translate poetry? To this there is only one possible reply – why not? (Delavenay 1960:109)

Admittedly, newspaper language also has its own distinctive style both with regard to syntax and semantics but that does not necessarily present the translator with any extra problems. Journalistic language exists within all languages whose culture includes newspapers and magazines. A machine could presumably be supplied with a specialised programme for the journalistic framework.

The emphasis here will be on problems at the semantic level since it seems that this is where most difficulty would arise for a non-human translator.

THE HUMAN TRANSLATOR

There is a radical difference between the way in which a translator deals with the language of the text he is working on and the way a listener processes the language he hears as part of a spoken dialogue. The translator is already in possession of the “whole” in advance: in other words he has a holistic overview of a

passage before he sets about the task of translating it. In a spoken dialogue the listener receives small chunks at a time and must wait for the following information before he can be certain of the message.

A certain amount of anticipation is also involved, but a listener can never predict with absolute certainty how a conversation will proceed:

All discourse, all interpretation of discourse works at a word-for-word and sentence-for-sentence level. There is no privileged access to underlying totality (Steiner 1975:294).

In terms of top-down and bottom-up processing the translator initially works top-down – from an overview of the whole text down to the details. The translator reads through the original text first thus:

1. whole passage (in order to get the gist)
2. smaller chunks, e.g. paragraphs
3. sentences
4. phrases
5. words

Before embarking on any kind of translation process where he will be working in the bottom-up direction, from the parts to the whole, the translator has already reduced the potential meanings of many lexical items and does not need to do a painstaking search amongst the possible meanings for each individual item. This is where a machine, without the advantage of a holistic view, would need to do some repetitive searches.

The question of consistency is also a very important one. Having translated a lexical item into the corresponding item in the target language the translator consistently uses the same translation throughout the text, unless he perceives that the meaning in one place differs and therefore requires a different translation. The machine, which is basically working at the word/sentence level, is not capable of making such decisions, but is, rather, seeing the same word for the first time every time it encounters it.

Once the translator sets about the actual task of translation bottom-up processing assumes a more important role. Nevertheless it is unlikely that a human translator ever looks at a lexical item in total isolation, even at a working level, whereas a machine inevitably does so.

Continuity is also part of the translator's skill. Whilst working on the details he constantly bears in mind the whole piece, so that he is, so to speak, working on two levels at the same time. In particular those parts which he has already translated are easily accessible and he does not need to constantly refer back to check

how a particular item was translated before. More importantly he can recognise where a new translation is required from the sense of the text.

A human translator also has the built-in ability to check his own work. He can spot his own mistakes in retrospect and correct them accordingly (allowing, of course, for human error). A machine on the other hand has no self-critical facility. Once a translation has been produced the task is complete and it has no device for monitoring its own output.

SYNTACTIC PROBLEMS

It is difficult to separate syntax and semantics in such an analysis as this. There are some general points which can be made however. Since syntactic patterns are a recurring feature in languages it is much easier to envisage a programme which could, for example, systematically translate word order differences. In particular two languages such as English and Swedish which have a basically similar word order are fairly compatible as long as the correspondence between them is consistent.

An example from the text under analysis here is the ability for Swedish to separate article and NP by the insertion of a qualifying phrase:

en, i ordets sanna bemärkelse, verkligt slumpmässig talsekvens
a genuinely random number sequence, in the true sense of the word

Written English does not allow this type of construction – the qualifying phrase is placed outside the ART+NP sequence. Similarly:

varje idag befintlig slumpvalsstrare
every random number producer in existence today

However where there is Art+Adj...NP the insertion of a qualifying phrase seems to be acceptable in English too:

ett visst, om än mycket otydligt mönster
a certain, though admittedly very vague, pattern

If this is a general correspondence for this particular word order sequence programming for it might not present too much difficulty. But there are other examples where the correspondence of word order pattern between the two languages is much less consistent, hence less programmable.

An example of this is the insertion of an adverbial phrase between *att...+ INF* in Swedish. Sometimes English allows this construction, other times not:

(Folk som ombeds) att slumpmässigt skriva ned
(People who have been asked) to write down at random

Whilst not an impossible construction English prefers to have the adverbial phrase after the verb. On the other hand:

för att därigenom garantera
to thereby guarantee

This is acceptable in English. Here it seems to be a question of style and a machine could not make such fine and arbitrary distinctions.

The passive

The passive construction in Swedish presents difficulties for translation into English for a number of reasons. One form of the passive is represented by the morpheme *-s* added to the end of the verb and is used quite frequently in Swedish. It presents two problems: firstly the passive is not used so frequently in English as a general rule and often has to make use of a paraphrase of the Swedish; secondly, even where the passive has a parallel use in both languages the structure in English is not the same and often requires that a whole sentence be restructured.

Ingen siffersekvens, som framställs genom en enkel dataprocess, kan sägas vara slumpmässig
No sequence of numbers produced by a simple computer process can be called random

När kortleken samlas in blandas den ofullständigt och delas ut igen
When the cards are gathered up they are insufficiently shuffled and then dealt again

In the first example above the present passive is used for the first two translations but the third use omits the word *are* because it has already been used previously in the same phrase. English often omits auxiliary verbs in this way, more as a matter of style than necessity, but a machine would have to treat every occurrence of the passive in the same way, since it cannot be sensitive to all of the differences in structure which English uses to translate the passive from Swedish. There are further problems involved with the passive which will be taken up later in the section on semantic problems.

The infinitive

The infinitive has already been mentioned above with regard to the fact that Swedish allows the insertion of an adverbial phrase between *att* + INF whereas English prefers it to be outside the sequence. As with the passive the infinitive is not always found in the same situations of use in the two languages. Often the Swedish infinitive is more naturally translated into English with a different structure – frequently the *-ing* form of the verb. The problem is then deciding where to translate into the English infinitive and where to use *-ing* instead.

Att blanda, skaka eller snurra...

Att veta...

Att blanda kort...

These examples are all translated more naturally by the *-ing* form of the verb e.g. *mixing, shaking* or *spinning*... where even the first use of *att* is not translated as the infinitive in English. We could perhaps provide the machine with a rule for this: where the infinitive is subject of its own clause translate Swedish infinitive with English *-ing*.

(Folk som ombeds) att slumpmässigt skriva ned
to write down at random

Here, on the other hand, the infinitive would also be required in English, but the machine would need instructions in order to recognise the different functions of the infinitive in Swedish and thus make the necessary decisions for translation.

Negation

Negation is a problem for translation at both levels – syntactic and semantic. English requires the auxiliary verb *do/don't* except with the verb *to be* and with modal verbs. In addition the verb must agree in person with the subject of its clause. Swedish has verb + negative particle in main clauses and negative + verb in subordinate clauses whereas English has a different pattern:

dessa procedurer tål inte
these procedures will/do not stand up to

att blanda kort är inte heller så lätt
mixing cards is not so easy either

att sex blandningar inte räcker
that six shuffles are not enough

då blandas den inte alls
then it is not mixed at all

As the examples from the text illustrate, the English negative marker prefers to come after an auxiliary or the verb *to be* but before a main verb in all types of clause. However it is not only the negative particle which is affected when translating – in the cases above where a negative phrase is used in Swedish a further restructuring of word order is also involved:

inte heller
not...either

(but cf. *inte det minsta...*
not in the least...)

inte alls...
not...at all

A further example from the text is in the title of the article:

Bli inte lurad av slumpen
Don't let chance fool you

Where the translation calls for the restructuring of a whole phrase (here, to provide the most natural translation for the imperative) the negative marker falls into its regular position within the target language structure.

The correspondence between English and Swedish negation is complex, depending not only on syntactic structure but also on the meaning of the negated phrase as a whole.

Tense

On the whole Swedish and English make a similar use of tense and similar structures for expressing tense are available in both languages. An exception is the English progressive aspect with *-ing* which has no direct equivalent structure in Swedish.¹

Therefore when the translator meets the simple form of the verb in Swedish there are two possible translations in English:

jag använder
I use *or* I am using

A machine would have to be provided with a method of determining where the progressive aspect is appropriate in English. In some situations rules can be formulated for the use of the progressive in English, but this is by no means always the case.

A further example of a difference of use of tense, in this case the future, comes from the text:

slumpen ska garantera
chance is supposed to guarantee

English does not use the future tense in this way so the translation will have to be with regard to the meaning of the original rather than following the structure.

SEMANTICS

Machine translation gets into deeper water when it comes to determining the meaning of individual lexical items or phrases so that they can be accurately translated into the target language.

¹ Swedish can of course express progressive aspect by using *hålla på med att* + verb, if it is required.

Occasionally what initially appear to be semantic problems can be resolved by syntactic analysis. Obviously these are at the phrase or sentence level. Ambiguity is one instance where a syntactic analysis may be able to disambiguate but this is not always the case.

There remain a lot of semantic problems, both for human and mechanical translators, since there is no tangible system of semantic meaning which could be used as an aid to translation.

It is probably most profitable to look at various areas of semantic problems rather than at individual lexical items, although these will also be used sometimes to illustrate specific problems.

Ambiguity

This has been much discussed in the literature on machine translation (Delavenay 1960, Wilks 1978, Sigurd 1986). In many cases, however, ambiguity is not a problem within the context of an actual text because many ambiguities are structural ones and if the machine analyses a sentence for syntactic functions a lot of apparent problems can be dispelled. If a word or phrase is ambiguous in the original text there is no reason why it should need to be disambiguated in the target text.

Most languages have homonyms – words which have the same form but differ in meaning. We are dealing here with language on the semantic level but in practice a translator never looks at a word in isolation on one level only. A simple example from the text in question is the Swedish word *att* which can be English *to* when it is before an infinitive, or the conjunction *that*. Obviously there are two different functions involved and this is reflected in the syntactic structure:

Man finner då (att sex blandningar inte räcker)
It is found that...

Folk som ombeds (att skriva ned)
People who have been asked to...

A machine should not have trouble recognising these different structures and therefore providing two different translations.

The problem becomes more acute when we look at those homonyms which have the same form and syntactic function but differ in meaning (polysemes). There are plenty of them in the text: *färg* 'colour, suit (pack of cards)', *kronor* 'crowns, heads (coin)', *blanda* 'mix, blend, shuffle (cards)'. The human translator has no difficulty in recognising that such words require a specific translation for the context they are in, because, as we have said earlier, he constantly has in mind the overall purpose and meaning of the text as he is working on it.

Perhaps the solution for a machine would be to furnish it with a set of *frames* (Minsky 1975) or *scenarios* (Sanford and Garrod 1981) which would consist of sets of lexical items for particular language situations. The human translator knows, for instance, that the general field of the text is games of chance, gambling, card playing etc. and he adjusts his view of the lexical items for translation accordingly. Such a solution would not be able to help, however, in a situation where the same word has different meanings within one text e.g. *rad* 'row, line, run, sequence'.

Compounds

The frame system mentioned above could also help get around the problem of compound lexical items whose translation is **not** a result of combining the translation of the parts e.g. *försöksråtta* (try rat) 'guinea-pig', *kortlek* (cardplay) 'pack of cards' (cf. *kortspel* 'card game'), *lekmannen* (playman) 'layman'. What we have to translate here is the meaning, not the individual words. This is a prime example of where word-for-word translation leads to confusing, sometimes even unintelligible mistakes. Malmberg 1986 discusses the possibility of a level of language below that of individual languages where meaning presumably converges. If such a level did exist and it could be encoded in some way it would of course make the work of all translation much easier. Malmberg himself doubts if such a thing exists, unfortunately: "a universal level of pan-linguistic structure which could serve as a point of departure for translations" (Malmberg 1986:20).

Prepositions

These are notoriously difficult to translate between most languages, since there rarely seems to be any neat pattern of correspondence. Swedish prepositions frequently cause trouble in English translation (and vice versa) because of the different interpretations possible. As a typical example we can examine the use of the preposition *på* in the Swedish text:

1. används på en mängd olika sätt
in a number of different ways
2. på begäran
at(on) request
3. kortfördelningen på
the dealing of
4. kommit på hur man
come up with

The fourth example above involves the use of a *particle verb* in Swedish, where the preposition is not separable from the verb. These are often difficult to translate – here, for instance, two prepositions are required to express the same idea in English. Even more problematical are cases such as the following (not in the text):

1. han [körde på] flickan verb + particle
2. han körde [på gatan] verb + adverbial phrase (P + NP)

These would surely be extremely difficult for a machine to translate since they are only distinguishable in spoken language by a difference in intonation (and possibly a pause). On paper they are not distinguishable except that the object of 1 is animate and that of 2 is inanimate. A human translator is therefore able to interpret correctly that 1 means that he ran the girl over (impl. in a car) and that 2 means that he drove along the street. Two entirely different translations are required to render the sense correctly in English.

Even for the more common uses of *på*, as for all prepositions, it is difficult to establish rules for their translation.

Delavenay 1960 demonstrates a possible elimination system for arriving at the correct translation of the preposition *of* from English into Russian.²

Ellipsis

By this is meant (here) translation situations where although a word in the original text is quite translatable it is not required and is generally omitted from the target language text. One particular lexical item which frequently occurs in the Swedish text is *som*. Depending on the situation it can mean 'who', 'which', or 'that'. When assigned the meaning 'who' as the subject of a clause English often omits it (together with the auxiliary verb):

- de som var födda
those (who were) born
- personer som genomgår
people (who are) undergoing

More important is the knowledge that as the object of the clause in the example below it may not be omitted:

- sådana som har skrivits ned
those which have been written down

² See Delavenay, page 58: Sub-routine for the determination of the exact meaning in Russian of the English preposition *of* (after Panov).

The question of omission is not a straightforward one, since tense is also involved.

Reference

Reference is an important feature of any text, serving to hold it together and making the reader perceive it as a text rather than a disconnected string of sentences (see Brown and Yule 1983, Chapter 6).

Referring expressions can be problematic, although the problem of reference is not as great for translation from Swedish into English as it is in the other direction. The Swedish language has more inflectional morphology than English, with marking for number and gender, so that referring expressions require a lot more checking for agreement than their English counterparts.

Pronouns always need to be treated with care in translation because they can refer backwards in a text (anaphors), forwards (cataphors) or outwards (exophors). *Det* and *den* occur frequently in the Swedish text and are both translatable as *it* in English. However, *det* also occurs in other expressions:

det var något fel på datorn
there was something wrong with the computer

It is quite likely that a machine would also translate *det* in this example as *it*, and likewise with *det finns* (there is), whereas *det gäller* on the other hand does require *it* (it's a question of).

Another type of referring expression is the possessive pronoun:

människan har fått betala ett högt pris för sin lättja

Sin refers back to *människan* and would be translated into English as *his* in order to agree with *man*, but it could equally be *her* or *its*. Access to semantic information is required to make this decision.

A third instance where reference within the text is important is in the use of reflexives. Swedish has a lot of reflexive verbs which have no reflexive counterparts in English:

matematiker tror sig ha kommit på
mathematicians believe they have

konsumenterna har hittills fått nöja sig
consumers have up to now had to be satisfied

använda sig av skiljer sig
to use differs

But there are other verbs which have a reflexive use in Swedish and require a reflexive pronoun in English too. These reflexive pronouns must agree with an

antecedent in number and gender and here again it is a potential problem for the machine, e.g.:

man frågar sig	hon frågar sig
one asks oneself	she asks herself etc.
(you ask yourself)	

Still thinking about *som* which was mentioned in the previous section, we can also see that there are potential problems in deciding where to use *who*, *which* or *that* in the English version of the text. The translation will depend on the previous reference according to whether it is human/non-human:

de datorer som ersätter
which (that)

kortspelare som blandar
who

Definite/indefinite

The correspondence between Swedish and English in their use of definiteness is not always consistent. There are a number of examples in the text where Swedish uses the definite article but English does not:

en av den moderna datavetenskapens
one of modern computer science's

bli inte lurad av slumpen
don't be fooled by chance

människan är en lat varelse
man is a lazy creature

Occasionally the situation is reversed:

om man singlar slant
if you toss a coin

löpte större risk
ran a greater risk

Here we really do have a situation where a feel for the language is the only reliable way of producing an accurate translation since there are few, if any, discernible rules which could be used as guidelines.

Morpheme -s

The *s* morpheme was mentioned earlier in connection with its function as a marker for the passive. However, this is not the only function it can have – it can also be reciprocal or be an active verb. There are certain, so called deponent verbs, whose only existing form is with the *-s* morpheme (see Jørgensen and

Svensson 1986:30). When it comes to translating these verbs care is needed since they require a variety of different forms in English.

tycks nu finnas
seems (deponent verb) to be (active)

kan sägas vara
can be said to be (passive)

statistiker/psykologer träffades
met together (reciprocal)

The translator is able to distinguish these different functions and translate them accordingly, but it is not certain that a machine would be able to recognise these differences as they only become obvious in context.

Cultural differences

Texts for translation are usually bound to a specific culture and every language inevitably makes reference to parts of its own culture. Even for the human translator specific cultural references can be impossible to translate, e.g. *Ole dole doff*. At best the translator can try to find an equivalent concept in the target language, otherwise it has to be left in its original form and an explanatory note added for the benefit of the reader.

Lexical gaps are also tied, indirectly, to the background culture of a language. Lyons (1977:301ff.) calls them *a hole in the pattern* because they are not simply the absence of a lexeme corresponding to the absence of a concept within a culture. There is, for instance, no equivalent word in English for *syskon* 'brothers and sisters'. The word *sibling* does exist but is definitely not in every day use in the way that *syskon* is.

An example of such a lexical gap in English is revealed by the word *man* which appears frequently in the text we have been examining. The word *one* certainly exists as a possible translation but by no means has the same kind of distribution as the Swedish word. More often English is likely to use second person pronoun *you* and sometimes avoids it altogether by using a passive or some other restructuring of the phrase:

hur man får fram...
a method...for obtaining

man finner då att...
it is found...that

eftersom man kan köra
since you can/since it is possible

att veta hur länge man ska blanda
knowing how long to mix

There are also lexical gaps in Swedish, where a whole phrase is most neatly translated by a single word in English:³

på varandra följande
consecutive

Idioms

There are sometimes certain expressions which, if directly translated, appear very odd to the native speaker, because an equivalent expression already exists as a given *chunk* in the language:

på sätt och vis
in a way

enligt konstens all regler
according to the book/by the book

tiden är dyrbar
time costs money

allt annat än
anything but

Translating the words as they stand, as a machine would do, will not produce an accurate translation.

The point being made here and in the previous section on culture is that a background knowledge of the cultures of both languages is essential to some aspects of translation work. This knowledge is of such an individual nature, varying greatly even from translator to translator that it could never be encoded in a system. In this case there is simply no substitute for a human translator.

CONCLUSION

One of the key words throughout this analysis has been context. The translator is never carrying out his work in a vacuum, whereas, essentially, this is precisely what a machine does.

The machine, at present, is in the same position as a foreign language learner who looks up words in a dictionary but is unable to make profitable use of what he finds because his knowledge of the context of use in the target language is so limited.

The greatest disadvantage of the machines in existence today is that they cannot *see* beyond sentence level (to some extent word level) and they are

³ Swedish has borrowed *konsekutiv* from English but it is not certain that this is in general use.

working on a small scale without having any *idea* what the text as a whole is about.

Another stumbling block to the machine is the fact that a human translator is, in the greater part of his work, translating the sense and not the individual words (see Malmberg 1986). The problem is, then, how to give the machine the ability to analyse the sense of an expression – an impossible task?

What we must remember is that the machine is an aid, a helping device, and not a replacement :

If translators are to co-exist with computers we must become actively involved in directing their uses, let us be their masters and they the tools. If we are unaware of, or fail in this respect, we have only ourselves to blame if, ultimately, we are the slaves of computers, compilers of word-lists and one-to-one glossaries, constrained to write in words and forms that machines understand. (Snell 1978, Introduction)

REFERENCES

- Brown, G. and G. Yule. 1983. *Discourse Analysis*. Cambridge.
- Delavenay, Émile. 1960. *An introduction to machine translation*. London: Thames and Hudson.
- Jørgensen, Nils and Jan Svensson. 1986. *Nusvensk grammatik*. Lund: Liber.
- Lyons, John. 1977. *Semantics I*. Cambridge: Cambridge University Press.
- Malmberg, Bertil. 1986. 'Arbitrary and motivated in language and translation – a theoretical introduction'. *Translation studies in Scandinavia*, ed. Lars Wollin and Hans Lindquist, 12-23. Lund: Gleerup.
- Minsky, Marvin. 1975. 'A framework for representing knowledge'. *The psychology of computer vision*, ed. Patrick H. Winston. New York: McGraw-Hill.
- Sanford, A.J. and S.C. Garrod. 1981. *Understanding written language*. Chichester.
- Sigurd, Bengt. 1986. 'Automatic translation – past, present, future.' *Translation studies in Scandinavia*, ed. Lars Wollin and Hans Lindquist, 51-60. Lund: Gleerup.
- Snell, B.M, ed. 1978. *Translating and the computer*. Amsterdam: North Holland.
- Steiner, George. 1975. *After Babel. Aspects of language and translation*. London: Oxford University Press.
- Wilks, Yorick. 1978. 'Machine translation and artificial intelligence'. *Translating and the computer*, ed. B.M. Snell, 27-43. Amsterdam: North Holland.

Lund University, Dept. of Linguistics
Working Papers 33 (1988), 65-76

Attitudes towards Varieties of Swedish

Hans Dahlbäck

Abstract

In this paper, results from a survey of attitudes towards regional and other Swedish language varieties are presented. Over 700 teenagers in the Swedish city of Malmö were asked about attitudes towards the Malmö dialect and other varieties of Swedish. Urban varieties were found to be most popular, immigrant Swedish least. The Malmö dialect was preferred in local media and Standard Swedish in national radio and television, although in entertainment both were accepted. Girls were more negative than boys towards the local dialect and more prone to code-switching.

INTRODUCTION

Speech variation continues to be an object of popular interest and evaluation, and attitudes towards speech varieties have in recent years become a concern for linguists as well. Another dimension of variation has been added by immigration to Western and Northern Europe, creating new and often negatively evaluated ways of speaking the language.

To study attitudes towards dialects and immigrant speech in Sweden, I have carried out a questionnaire survey among over 700 teenagers in Malmö. The study has the twofold purpose of a) investigating the evaluation of different varieties and venture some generalizations, and b) exploring to what extent sex and socio-economic status of the individual influence attitudes.

After some preliminary sections dealing with earlier research on language attitudes in Sweden and elsewhere and about regional variation in Sweden, the survey is described and the results are given.

ATTITUDES TOWARDS VARIETIES: SOME POSSIBLE DETERMINANTS

There are by now a great number of studies that report differential attitudes towards Standard and Non-standard speech, even if it is the same person speaking in both cases, as in the so-called matched guise experiments. Generally, speakers of the Standard are rated highly in terms of social competence factors such as intelligence, whereas Non-standard varieties confer to the speaker characteristics of social attractiveness, e.g. friendliness, sociability and toughness. Cf. Trudgill 1983.