THE STRONG AND THE WEAK POINTS OF TEXTS TRANSLATED BY MACHINE IN COMPARISON WITH TEXTS TRANSLATED BY HUMANS

Merja Ågren

CONTENTS

1. INTRODUCTION

A machine translation (MT) system is a software that automatically translates a text from one language into another.

First, in the section 'Theoretical background', I will discuss briefly
the kinds of problems that an MT system encounters as these problems
differ from the problems of human translation. The theories that are
used include only those of machine translation and not translation in
general. This is because the aim of the study is to discover the strong
and the weak areas in MT and to find the areas in which machine
translation could be helpful to human translators.

Secondly, I will give a short review on the history of machine
translation outside of Finland as well as in Finland. Today there are
several commercially available MT systems. Unfortunately, the systems
that have been developed outside of Finland normally do not include
Finnish. In Finland, Kielikone Oy has developed a machine translation
system called TranSmart which translates texts from Finnish into
English. The translations that I will be using in the 'Analysis' are
produced by this MT system. I will also describe the situations in which
MT is used nowadays and the situations in which it will be used in the
future.

In 'Analysis' I will analyse the English translations of five Finnish
texts. The texts are translated by TranSmart and by first and second
year students of English in the University of Oulu. I will categorise
the errors that I can find in those translations and draw a summary of
the errors. I believe that human translators and the MT system will make
different types of errors and hope to find some areas in which machine
translation can be of help for human translators.

2. THEORETICAL BACKGROUND

In order to understand the state of machine translation today one has to
know something about its past. After centuries of work by numerous
scientists the dream of the 17th century is only now becoming reality.
However, no single, flawless solution has been found to tackle all the
problems of MT. Consequently, even today there are several ways to
perform the task. One successful approach has been developed by
Kielikone Oy. I will discuss the translations produced by their MT
system later in this paper.

2.1. Difficulties in machine translation

During the translation process a machine translation system encounters
many problems whose solutions seem self-evident to a human translator.
The reason is that the machine approaches translation differently than a
human. A human reads the text to be translated first and figures out the
meaning of it and then translates it. Because it is difficult for the
machine to cope with meanings it has to base its translation on the
structures of language (Arnola 1995: 42).

2.1.1. Levels of ambiguity

Scott (1993,174-175) divides the difficulties that a MT system faces
into five levels:

On the first level, the lexical syntactic level, the system searches a
definition for the terms from its dictionary.

If there are more than one definition the system has to go to the second
level, sentential syntactic level, where the term is placed into its
context. However, at this stage the system does not consider semantics

but only sees the sentence as a syntactic string containing elements that have more than one possible interpretation.

The third level, lexical semantic level, consists of semantic-property codes that deal with ambiguities such as whether in fig. 1. 'clean' modifies 'sheets' or both 'sheets' and 'blankets'. It also handles multiple meanings of words (in fig. 1. the meanings of 'keep').

The fourth level, sentential semantic level, deals with the semantics of the sentence in question. For example, it translates the meanings of prepositions.

In order to get into the fifth level, extra-sentential level, a MT system should be able to function in the level of discourse, that is to understand things such as ellipses and anaphora. Scott (1993: 175) admits that this is an extremely hard task for a system and argues that the semantics of the systems need to be improved. Arnola (as cited in The Finnish Formula 1994: 5), however, is sceptical of this of this approach. He argues that well-designed syntactic processing of the sentences is sufficient. He says: "Semantics is a swamp. It brings with it a raft of difficulties: data representation problems and great complexity problems." Nevertheless, he admits that the dependency model (see fig. 2.) in which the Kielikone system is based on has a "flavour" of semantics.

Different MT systems cope differently with the problems in the levels of ambiguity. Some systems handle only the first two levels, other more sophisticated systems tackle the difficulties of the deeper levels with varying success. The levels of ambiguity are illustrated in fig.1.

Fig. 1. The five levels of ambiguity (Scott 1993: 175).

2.1.2. Linguistic theory

Melby (1996: 93-94) argues that it is difficult to apply any linguistic theory on machine translation because machines use a different process than humans when translating (cf. Arnola 1995: 42). According to him the fact that MT is able to produce high-quality translations of controlled language (i.e. of a certain sublanguage) but not of natural language suggests that "current mainstream linguistics may not have captured the essence of natural language beyond the realms of morphology and syntax". He believes that applied linguistics could further studies in linguistics more than vice versa.

2.2. On the history of machine translation

As early as in the 1600's researchers tried to invent mechanical devices that would translate texts from one language to another. They argued that there was a need for a universal language for natural languages were inadequate in scientific communication because they were too ambiguous. As a result, they proposed that numerical codes could be used to mediate among languages. Their idea of giving each lexical element the same code number in all languages lead to the compiling of several such mechanical dictionaries in the middle of the seventeenth century. In the 1950's and in the 1960's when MT research was very active some researchers found that these attempts to create a universal language

were the true pioneer work in machine translation (Hutchins 1986: 21-24).

The electronic digital computer was invented in the 1940's. Machine translation was one of its first linguistic applications. The expectations for this application were high. In fact, many of the early researchers believed that computers would provide the solution for translation problems all over the world (Simons 1984: 154, 168).

In the 1950's when tension grew between the USA and the Soviet Union and the cold war began there was a great demand for English-Russian MT systems. Americans believed that if they were ahead Russians in technology, they would secure USA's position as the leading world power. Therefore, the government eagerly financed several MT research projects. Unfortunately, many of the researchers viewed translation as a relatively simple, straightforward task: they believed that combining a bilingual dictionary and grammars of the languages was sufficient. They did not take into account that there is only a thin line between syntax and semantics and consequently they did not even try to analyse the structure or the meaning of the sentence to be translated. The result of these projects was a number of unreliable MT systems. In the 1960's most of the projects were terminated as unsuccessful and in the USA a period of over ten years followed during which there was hardly any research on MT (see, e.g. Simons 1984: 154, 168; Arnola 1995: 41).

Arnola (1995:42) and Hovy (1993:167-168) discuss the state of machine translation today. MT research became fashionable again in the 1980's. Japanese industry needed something to lower the language barrier on its exports. Also the European Community was looking for ways to lessen the enormous workload in translation of documents. Instead of using direct MT systems, which translate texts by substituting the source language (SL) text word by word or a short phrase at the time with the words and phrases of the target language (TL) , now the researchers agreed that the sentences to be translated needed some kind of preliminary treatment before the actual process of translation. What that treatment should be, however, was a matter of dispute.

2.2.1. Interlingua systems

Some researchers believe that the interlingua technique is the solution. This ambitious technique attempts to open the meaning of the source language sentence by creating a representation of it that is midway between all languages and then to print the meaning in the target language. This system is useful because it requires less transfer rules (i.e. rules that define how a certain thing is translated) than the transfer systems but it needs to make a more careful analysis of the source language sentence. However, opening the meaning of a sentence is such trivial task that no adequately functioning interlingua systems have been developed (Arnola 1995:42; Hovy 1993:172,175).

2.2.2. Transfer systems

A more popular opinion among the researchers is the support of the transfer technique. As most modern MT systems, also the Kielikone system is based on the transfer technique. In the transfer technique the translation is based solely on structure, and not on the meaning, of the source language sentences because, according to the supporters of this view, it is only the structures that can be studied reliably and extensively (Arnola 1995:42).

Like all systems, also transfer has its disadvantages. It needs transfer rules between all languages it translates. In a case of a system which includes three languages that can all function both as a source language and as a target language there has to be six sets of transfer rules - between languages A and B, A and C, and B and C in both directions. If a new language is introduced to the system new rules have to be created with all the three other languages and therefore the number of the sets of rules is 12. When the lexicons and the grammars that are included in the systems are extensive, it is not surprising that creating a system that translates numerous languages is exhausting (Hovy 1993: 175).

2.3. On famous machine translation projects

Today there are several more or less successful commercial machine translation systems. However, numerous projects have also failed. I do not intend to discuss the available systems nor the failures in detail but only review shortly the two probably the most famous ones.

2.3.1. Systran

According to Sager (1993: 273) Systran is the oldest and one of the most successful commercially available MT system. It began as a direct MT system but later developed into a transfer system (Hovy 1993: 171). Today it offers 27 language pairs and according to Miller (1993: 182-183) among non-Finnish commercial MT software, it is the only one that offers Finnish as a target language. None of the systems that are developed outside Finland have Finnish as a source language.

2.3.2. Eurotra

Eurotra was a transfer-based MT project financed by the European Community. It was launched in 1979 and it was hoped to solve all translation problems around the community.

The system included nine languages and 12 countries. 200 people worked in the project and the budget was 200 million Finnish marks. Consequently, the project was too ambitious, too massive and too dispersed. It was terminated as a failure in 1992 (Sager 1993: 273, Keinonen 1995: 35).

2.4. Machine translation with the Finnish language

In Finland machine translation with the language pair Finnish-English has been studied in two research groups. In the University of Helsinki a group lead by Professor Lauri Carlsson has studied machine translation from English into Finnish. Kielikone Oy, on the other hand, has developed a computer program that translates texts from Finnish into English (Arnola 1995:44). Since the translated texts that I will examine later in this paper are produced by the Kielikone system, TranSmart, I will discuss only the work of this group rather than the other.

Naturally, research has been done with other language pairs, too. The Finnish Meteorological Institute, for example, uses machine translation in translating weather forecasts from Finnish to Swedish. The use of this MT system, however, is limited to the sublanguage of weather forecasts (Blåberg 1991).

2.4.1. Kielikone Oy

Suomen itsenäisyyden juhlarahasto ,SITRA, started a project called Kielikone in 1982 to develop computational models for Finnish. Lack of those computational models was one of the main reasons why Finnish companies were not able to develop machine translation at the time. The aim of the project was to develop language technology products and at the same time create an infrastructure for MT research. The project has produced several software products for Finnish e.g. a morphological analyser and spelling checkers based on that model, a morphological synthesiser, a hyphenation algorithm, dependency parsers and a synonym dictionary. It has also produced an electronic bilingual Finnish-English dictionary (Jäppinen et al. 1993: 173; Jäppinen et al. 1991: 107).

In 1987 the Kielikone research group found that there was a need for a MT system among large Finnish companies. The focus of research became machine translation and Kielikone, the research project, turned into Kielikone Oy, the company (The Finnish Formula 1994: 4).

In 1992 the MT project reached the product development phase and in 1993 the first customers started to use TranSmart in their technical documentation. Kielikone's first customer was Nokia Telecommunications Oy but Trantex Oy and Rautaruukki Oy soon followed(The Finnish Formula 1994: 4, Nenonen 1995: 64).

2.4.2. TranSmart

The Kielikone system, TranSmart, is based on the transfer technique. It produces a raw translation from Finnish into English that sometimes needs to be edited by a professional translator. The amount of post-editing depends on the level of quality that is required from the translation. In description of the process of developing the MT machine in Kielikone project Jäppinen et al. (1993) states that the machine is a tree manipulation system which produces dependency trees (F-trees). The nodes of the tree represent the words and the branches represent binary dependency relations between the words of a sentence (fig.2).

Arnola (1995, 43) describes the actual process of translation that is done in three phases:

        1. Analysis

        First the parser, which is the most important part of a machine translation system, analyses the relations of the words in the sentence. It produces the kind of dependency tree as is presented in fig.2. In a sentence "Pitkällisen riitelyn jälkeen he istuivat alas", for example, the parser has to decide that 'jälkeen' is a postposition and not a form of the noun 'jälki' and also that it is an adverbial for the verb 'istuivat' (see fig.3.).

        2.Transfer

        In the transfer phase the MT system searches the target language equivalents for source language words and expressions in its dictionaries.

        3. Synthesis

The sentence reaches its final form in synthesis when the necessary prepositions and negatives are added. Synthesis also generates plural forms for the nouns and the correct forms for the verbs.

The whole process with the representation of the dependency trees is presented in fig. 3.

Fig.3. The machine translation process (Arnola 1995, 43).

2.5. The use of Machine translation

Arnola (as cited in Nenonen 1995: 64) stresses that machine translation is not suitable for all texts. Juridical texts, contracts and literature need to be translated by a human translator. He also says that if a high-quality translation is required the work must be done by a professional, too.

Melby (1996: 89-99) introduces three conditions which have to be met if MT system is expected to produce a high-quality output:

   1.The texts to be translated must be restricted to a well-defined domain of knowledge shared between source and target languages.
   2.The source texts must be carefully controlled to conform to a formal syntax and semantics. Such texts are said to be in a "controlled language."
   3.The machine-translation system must be tailored to the domain and the controlled language.

He adds that under these conditions less than five percent of texts being translated are high-quality machine translations. It is due to the restrictions of machine translation that it does not compete with professional translators. Human translators often find the translating of "controlled texts" or restricted texts uninteresting. So, in fact, in these situations nowadays many translators use MT to help with their work.

Machine translation is suitable for the translation of technical texts, news, patent applications and patent abstracts. The best feature of MT is its speed. Furthermore, in many technical documentation projects there is vast amount of text and many translators working as a team. It has to be made sure that each translator uses the same terms for a particular function. This problem can be avoided by using MT (Nenonen 1995: 64; Keinonen 1995: 34).

One interesting application of MT is e-mail communication. Because e-mail is confidential it is impossible to have a human translator "on-line". Other Internet applications also seem promising. Arnola (as cited in Keinonen 1995: 34-35; Nenonen 1995: 64-65) also has many other visions on the future of machine translation as well. It could be used in fax-machines and Xerox machines. Maybe sometimes after the year 2005 we will also have a simultaneously translating telephone.

3. ANALYSIS

The purpose of my study is to compare texts translated by a machine translation system to texts translated by a human translator. My aim was not to show whether one was better than the other (as I have already pointed out, today machine translation and human translation are rather

synenergic than competitive), but merely to illustrate some of the strong and the weak points of both the ways for producing translations.

3.1. Material

Kielikone Oy supplied me with the Finnish originals and the English translations produced by TranSmart of five texts. I asked ten first and second year English students in the university of Oulu to translate the same texts. I chose to have first and second year English students rather than a professional translator or a person with only lukio level knowledge of English as the human translator in order to find out "the skill level" of the MT system. The quality of machine translation today certainly is so good that the translations it produces are better than translations produced by a person with only an "average" command of the English language. However, quite often the translations are edited by a human. Therefore, I did not want to compare the translations with the work of a professional translator, either, because as mentioned above, the aim of my study was not to determine whether one way of producing translations is better than the other but simply to compare the errors made by both the machine and the human and, if possible, to find some areas in which MT would be helpful for human translators.

Originally I wanted to compare texts from various genres but Kielikone Oy informed me that the MT system is not designed to translate literature because the machine does not understand stylistic differences. Therefore, the five selected texts were news articles from Finnish newspapers and magazines.

3.2. Method

I checked the translations and marked the errors in them as well as I could. The percentages may vary by +/- 2 or 3 depending on how strictly the translations are marked and due to the fact that I, too, am a foreign learner of English. Nevertheless, I believe that the figures are indicative of the number of different types of errors. In most cases, I chose to ignore possible stylistic errors both in machine translation and in human translation partly because of the machine's inability to recognise stylistic differences, partly because stylistic differences are often a matter of taste and therefore it is often hard, if not impossible, to say whether something is really an error or only a clumsy expression.

I divided the errors into eight categories:

   1.Vocabulary. This category includes all word classes, except articles, prepositions and pronouns. The word causing an error in this category was either a totally wrong word in its meaning or it was right in its meaning but it was used in the wrong context.
   2.Articles. Using the wrong article, absence of an article or using an article where it is not needed are all counted as errors in this category.
   3.Prepositions. Inaccurate uses of prepositions are included in this category.
   4.Pronouns. This category includes all pronoun classes, although the pronouns causing most of the troubles seem to be the personal pronouns and the possessive pronouns.
   5.Tenses. Using the wrong tense is counted as an error in this category. It also includes errors within tenses in such instances as whether to use was or were or whether to use the present tense with or without the third person singular '-s'.

6.Clause structure. Sentences that did not follow the conventional English clause structure were included in this category.

7.Paraphrase. If the translated sentence was only a paraphrase of the original sentence (i.e. it only vaguely describes the meaning of the original sentence) I have counted it as an error. It could be argued that this is semantic translation but I have chosen to follow the guidelines used in translation courses in the University of Oulu. In those courses paraphrase is not an acceptable method of translation.

8.Inadvertence. This category naturally applies only to human translations. The errors in it are missing words or sentences and other such mistakes that result from misreading the source text, not from lack of competence.

The examples given do not cover all the mistakes in a category but I tried to choose them so that they would illustrate the most common mistakes in that particular category. They may also contain other errors than the underlined one, but I have drawn attention only to those mistakes which are relevant within a category.

3.3. Analysis of the human translations

3.3.1. Vocabulary

This category seems to have been the most problematic one for human translators. 44% of all errors fall into this category. It is impossible for a human to memorise the whole content of a dictionary or dictionaries. Therefore, even if one has a good bilingual dictionary it can be difficult to find the right expression if the definitions for the TL terms do not seem to match completely those of the SL terms. A good English dictionary is not of help either if one has no idea of what term to look up. More commonly, however, the vocabulary mistakes seemed to result from negligence. Some of the human translators had not bothered or did not have the time to check whether a term they were not familiar with was correct in the particular context or, indeed, the correct term in any context.

(1) incorrect term
SL: Kokaiini oli yhä alkuperäisessä pahvilaatikossa, jossa oli joskus ollut kolumbialaisia tietosanakirjoja.
TL: The cocaine was still in the original cardboard box which had sometimes contained Colombian dictionaries.

According to Sinclair (Collins Cobuild English Dictionary 1995) and Hurme et al.(Uusi englanti-suomi suursanakirja 1990), an encyclop[a]edia is "a book or set of books in which facts about many different subjects or about one particular subject are arranged for reference" = tietosanakirja where as a dictionary is 1. "a book in which the words and phrases of a language are listed alphabetically, together with their meanings or their translations", 2. "an alphabetically ordered reference book on one particular subject or limited group of subjects" = sanakirja. Even the second definition of dictionary does not imply that a 'dictionary' is as comprehensive book as 'tietosanakirja'.

Some of the errors resembled those of inadvertence resulting from misreading the text or from a typing error.

(2) incorrect term (caused by inadvertence?)
SL: Virallisesti kiertueen tarkoituksena on antaa tukea entisten sosialistimaiden demokratioille.

TL: Officially the purpose of the tour is to give support to the democracies of the formal socialist countries.

3.3.2. Articles

Only 3% of the errors made by humans fall into the category of articles. There were cases where the definite article was used instead of the indefinite article although the noun was not mentioned before and it could not "be identified uniquely in the contextual or general knowledge shared by speaker and hearer" (Greenbaum & Quirk 1990: 77) and vice versa. Furthermore, there were errors with the indefinite article a/an:

(3) the indefinite article
SL: Ohjelmaan kuuluu todennäköisesti tutustuminen Helsinkiin.
TL: The program probably includes a introductory to Helsinki.

3.3.3. Prepositions

Errors in using prepositions were also 3% of the total number of errors made by humans. I found the number surprisingly low because usually the prepositions are hard to grasp for a Finnish learner. In addition, some of the instances included as errors in this category are in the borderline between being clear errors and being only awkward expressions.

(4) preposition
SL: Brittitutkijoiden havainnot eräällä Skotlannin rannikon saarella kumoavat tämän käsityksen.
TL: The discoveries of some British researchers on an island on the coast of Scotland disprove this view.

Logically thinking the island is not 'on' the coast but 'off' the coast. The Finnish is able to avoid the problem by using the genitive.

3.3.4. Pronouns

There were no errors belonging under the category of pronouns in the human translations.

3.3.5. Tenses

Incorrect use of tenses caused 14% of the errors. The future tense seemed to be most problematic. Here, however, it must be noted that one of the texts dealt with future events but the Finnish original, naturally, was written with the present tense. This text was translated by three of the students, two of whom used the present tense also in their translations. This continued throughout their translations and since each occurrence of this mistake was counted as an error, this increased the total number of errors in this category considerably. Other less serious errors were caused for example by whether or not to use the continuous form of a particular tense or by the use the present tense with third person singular subject without the ending '-s'.

(5) the future
SL: Rouva Clinton tulee Suomeen suoraan Itä-Euroopan kiertueeltaan.
TL: Mrs. Clinton comes to Finland straight from her tour in Eastern Europe.

3.3.6. Clause structure

Using a clause structure which is foreign to English syntax caused 8% of the errors in the texts translated by humans.

(6) clause structure
SL: Iltapäivän rouva Clinton viettää yksityisesti suurlähettiläsparin kanssa.
TL: The afternoon Mrs. Clinton will spend privately with the ambassadorial couple.

'Spend' is a transitive verb which requires an object, in this case 'the afternoon'. The normal structure for sentences like this is SVO, not OSV, as in the example, which is unknown sentence structure in English (Greenbaum & Quirk 1990: 204-207). This example is a quite typical error in this category, although not all of the errors were as evident as this one.

3.3.7. Paraphrase

Only human translators made these kind of errors. This is perfectly understandably because a human figures out the meaning of the text first and then translates it but the machine does not understand meanings. Therefore, it translates faithfully the words and the structures of the source text.(Arnola 1995: 42).

(7) paraphrase
SL: Kerran yksi joutsenpariskunta kiristi aina vähän väliä rinnalle ja ihmetteli, mikä vekotin vedessä oikein kulkee.
TL: Once a swan couple kept catching up with boat, wondering what it was.

The meaning of the TL sentence is the same as in the SL sentence but the style of the original is lost. Admittedly, these type of sentences produced by the humans were more fluent than those produced by the machine but that does not necessarily make paraphrase a valid translation technique (cf 3.4.6. example (18)). 8% of the errors made by human translators were caused by paraphrase.

3.3.8. Inadvertence

As the phrase goes to err is human. This seems to be valid also in concerning translation. 20% of errors made by humans were caused by inadvertence.

(8) missing sentence
SL: Tavallisen soutuveneen perässä on pikkuinen moottori. Vene liukuu vedessä kuin unelma, tasaisesti, äänettömästi ja saasteettomasti. Sähkömoottori saa voimansa auringosta. Suomalaisen soutuveneen on työstänyt aurinkoveneeksi Juha Nyman Särkisalosta.
TL: There is a small engine in the rear of an ordinary rowboat. The boat glides through water like a dream, smoothly, noiselessly and unpollutingly. _____ The Finnish rowboat has been developed into a solar boat by Juha Nyman from Särkisalo

(9) missing superlative
SL: Bill Amos kollegoineen sai DNA-analyyseillä selville, että vaikka rannalla pullistelevat voimakkaimmat koiraat keräävät ympärilleen haaremin, useimmat naaraat livahtavat muualle lempimään.
TL: Bill Amos with his colleagues found out through DNA-analysis that although the strong males, buffing on the shore, gather a harem around them, most females sneak away to mate.

The examples clearly illustrate that at the university level these kinds of mistakes result only from misreading the text.

3.4. Analysis of the machine translations

The examples given show only extracts of machine translated texts. To provide a more comprehensive view I have enclosed the Finnish original and the machine translation of one of the texts in the appendix.

3.4.1. Vocabulary

20% of the errors in the machine translated texts could be included in the category of vocabulary. Although 20% is a quite large portion of the errors, this was not the largest category.

(10) incorrect term
SL: Jos nykyisen veneen nostaa aina vesireissun jälkeen telille, puu kuivuu ja seuraavassa vesillelaskussa vene täyttyy vedellä.
TL: If one lifts the present boat always after the boat trip to the spindle, the tree will dry and in the following launch the boat will become full of water.

The wood that is used in the boat, in all likelihood, is no longer a living tree.

3.4.2. Articles

Incorrect use of articles caused 18% of the errors in the machine translations.

(11) the definite article
SL: Hasikset oli pakattu joko muovipusseihin tai folioon.
TL: The hashish had been packed either in the plastic bags or in the foil.

Neither 'plastic bags' nor 'foil' had been mentioned previously in the text (cf. 3.3.2.). In some cases the article was completely missing or the indefinite article was used instead of the definite article:

(12) missing article, incorrect article
SL: Ahtisaaret tarjoavat vieraalle lounaan Mäntyniemessä.
TL: __ Ahtisaaris entertain a guest to lunch in Mäntyniemi.

'The Ahtisaaris' refers to both the president Martti Ahtisaari and Mrs.Ahtisaari, therefore, the definite article is required (Greenbaum & Quirk 1990: 88). The 'guest' refers to Hillary Clinton and it is mentioned already earlier in the text that she will be meeting the Ahtisaaris.

3.4.3. Prepositions

Errors in the use of prepositions were 6% of the total number of errors.

(13) preposition
SL: Ihmisten lisäksi muutamat linnut tuntuvat olevan kovasti kiinnostuneita siitä.
TL: In addition to the people, there seem to be a few birds very interested of it.

The word 'interested' requires the preposition 'in' (Sinclair 1995). The other errors in prepositions were mostly similar to this one.

3.4.4. Pronouns

Arnola (1995: 43) states that personal pronouns are one of the most common causes for problems in machine translation, because the machine is not able to tell whether the subject is animate or inanimate. Indeed, there were quite many errors of this type in the translations. Furthermore, there were errors in whether the third person singular pronoun should be he or she. The errors in pronouns caused 18% of the errors in the machine translated texts.

(14) inanimate/animate
SL: Bill Amos kollegoineen sai DNA-analyyseillä selville, että...
TL: Bill Amos with its colleagues found out with the **DNA** analyses that...

(15) third person singular
SL: 1970- luvulla Hillary Clinton osti paljon Marimekon vaatteita ja on ollut edelleen kiinnostunut yrityksen kuulumisista. Tästä syystä hänen on veikattu pistäytyvän Marimekon myymälässä.
TL: In the 1970's %%Hillary%% Clinton bought many clothes of Marimekko and has been still interested in the news of the company. For this reason it has been guessed that he will drop in the shop of Marimekko.

If the MT system is not able to translate a word or if it is able to translate a compound only in parts, it marks these words with **...**, %%...%%, or *%...%* (cf. 3.4.6. example (18)).

3.4.5. Tenses

14% of the errors resulted from incorrect tenses. The future tense was problematic for the machine as well, although not as frequently as for the human translators. The simple past tense caused an another problem:

(16) the simple past
SL: Puolen tunnin hakkailun jälkeen miehet arvioivat, että huumeita heitettiin kuiluun liikaa. He onkivat neljä laatikkoa pois.

TL: After the flirting of half an hour the men estimate_ that too much drugs were thrown into the gulf. They angle_ four boxes away.

The whole text in question is written in the past tense. In Finnish, however, both 'he arvioivat' and 'he onkivat' have the same form in present and in the simple past. Therefore, it is hard to determine which tense to use in TL text simply by looking at the verb phrases. The machine translates one sentence at the time starting from the first sentence, so it is not able to figure out the right tense by comparing the sentence into the other sentences in the text. Even if the machine took the tenses in the previous sentences into account, in this case it would not have helped because the above sentences were the first sentences of the text.

3.4.6. Clause structure

This was the most problematic category for the machine to translate.

(17) clause structure
SL: Kaikki syytteet perustuivat pääosin miehen omaan kertomukseen.

TL: All the charges were based on mainly the man's own story.

When an adverbial is in a medial position it is normally placed immediately after the operator or the copula (Greenbaum & Quirk 1990: 161-162). In this case 'mainly' should be placed after 'were'. There were quite many sentences with errors like this where the English clause structure seemed to imitate the Finnish one.

In the machine translated texts, there were also a few sentences in which the clause structure seemed very strange. I assume that when translating these sentences it would have been essential to understand their meaning in Finnish for these were the same sentences that had the tendency to become paraphrased in the translations made by the students. As mentioned above, due to the nature of MT, it does not paraphrase sentences. Instead, these difficult sentences are translated faithfully which in some cases leads to awkward sentences:

(18) awkward translations
SL: Kerran yksi joutsenpariskunta kiristi aina vähän väliä rinnalle ja ihmetteli, mikä vekotin vedessä oikein kulkee.
TL: Once one *%swan couple%* tightened every little while beside and was surprised what gadget in water right goes.

All in all, these errors were 24% of the total number of errors.

3.4.7. Paraphrase

As mentioned earlier, due to the nature of MT, it does not paraphrase sentences.

3.4.8. Inadvertence

The machine is not influenced by human factors such as fatigue or interruptions. Therefore, none of the errors made by MT were classified as inadvertence.

3.5. Machine translation versus human translation

Table 1 presents a summary of the errors.

Table 1. Percentage of the errors in the translations

CATEGORY	HUMAN TRANSLATIONS	MACHINE TRANSLATIONS
1. vocabulary	44 %	20 %
2. articles	3 %	18 %
3. prepositions	3 %	6 %
4. pronouns	0 %	18 %
5. tenses	14 %	14 %

```
        6. clause structure
                                            8 %
                                                                    24 %
        7. paraphrase
                                            8 %
                                                                     0 %
        8. inadvertence
                                           20 %
                                                                     0 %
```

The study supported the idea that MT should not be seen as a
competitioner for human translators but, in fact, it can be great help
for them. More than half of the errors in vocabulary could be eliminated
by using MT, not to mention all the errors caused by inadvertence. Much
of the time the human translators use in completing a translation task
is spent in searching the correct terms from a dictionary. This is one
reason why machine translation facilitates the translation process.

The MT system made quite many errors in using pronouns but these errors
were easy to notice. The human translators, on the other hand, did not
make any errors in this category. This seems to suggest that post-
editing the pronouns in a machine translated text should not be very
difficult for human translators. At least to me, some of the errors in
prepositions and articles were harder to notice. Some of them, however,
were very clear.

In the use of correct tenses, machine translation did not seem very
helpful. Tenses caused the same amount of errors in the human
translations and in the machine translations. The future tense caused
most of the problems for both ways of translating.

Clause structure and paraphrase are more complicated cases. In sentences
that are difficult to translate the human translators seemed to have a
tendency to convey only the meaning of the SL sentence by paraphrasing
it in the TL sentence. In the same situations the MT system produced
more or less confusing sentences. Some of these awkward translations are
not easy to rephrase but luckily they were only a small portion of all
the errors in this category. Most of the errors that occurred in the
clause structure of the machine translated texts were quite easily
corrected.

There is one thing that gives MT an advantage over human translation
that did not show in my study. That is the speed of MT (Arnola 1995:
43). At the moment TranSmart is able to translate fifteen sentences in a
second. At least for me, it was faster to read and correct the machine
translated texts than it would have been to translate the texts
completely. Naturally, the amount of post-editing time that is required
depends on how high the quality of the translation has to be.

4. CONCLUSION

In my study I have briefly discussed the history and the problems of
machine translation. I have analysed and compared the translation errors
in five texts that were translated from Finnish into English by an MT
system called TranSmart and by first and second year English students. I
have tried to find out the strong and the weak points of machine
translation. In other words, I have determined some areas in which MT
could be helpful for human translators and some areas in the machine

translated texts which need to be post-edited by a professional translator.

Machine translation has developed tremendously over the decades. It has advanced from being numerical codes that mediate between languages, trough direct MT systems, into the present transfer systems. Whether it is possible to develop a fully functioning interlingua system still remains uncertain. One of the dreams that the early researchers had, however, still remains unaccomplished; MT has not solved translation problems world-wide. Even if (when) MT becomes more common as a translator and as a translation aid, the human translators need not worry that their jobs will disappear. After all, MT has increased the number of texts that are being translated and at least for now some of the MT texts have to be post-edited by professionals and not all translation can be performed by machines.

BIBLIOGRAPHY

Arnola, H. 1995. Konekäännin murtaa kielimuuria. Tiede 2000 6/95, 40-44.

Blåberg, O. 1991. On machine translation of Finnish weather forecasts. In Erikoiskielet ja käännösteoria.
VAKKI-seminaari XI. Vöyri 9.-10.2.1991, 22-27. Vaasa: Vaasan yliopisto. Kielten laitos. Käännösteorian ja ammattikielten tutkijaryhmä.

Greenbaum, S. & R. Quirk 1990. A Student's Grammar of the English Language. London: Longman.

Hovy, E. 1993. How MT works. Byte 18, 167-172, 175-176.

Hurme, R.., M. Pesonen, O. Syväoja 1990. Englanti-suomi suursanakirja. Porvoo: WSOY.

Hutchins, W. J. 1986. Machine Translation: Past, Present, Future. Chichester: Ellis Horwood Limited.

Jäppinen, H., K. Hartonen, L. Kulikov, A. Nykänen & A. Ylä-Rotiala 1993. Kielikone Machine Translation Workstation. In Nirenburg, S. (ed.), Progress in Machine Translation, 173-184. Oxford: IOS Press.

Jäppinen, H., L. Kulikov & A. Ylä-Rotiala 1991. Kielikone Machine Translation Workstation. In Proceedings of MT Summit III, 107-111. Washington D.C.

Keinonen, M. 1995. Konekäännöstä ihmisen avuksi. Tekniikan Näköalat 2, 34-35.

Melby, A. 1996. Machine Translation and Other Translation Technologies. In Grabe (ed.), Annual Review of Applied Linguistics, 86-98. Cambridge: Cambridge University Press.

Miller, L. C. 1993. Babelware for the desktop. Byte 18, 177-183.

Nenonen, H. 1995. Konekääntäminen on jo tuote. Kauppalehti Optio 6, 64-66.

Sager, J.C. 1993. Language Engineering and Translation: Consequences of automation. Amsterdam: John Benjamins Publishing Company.

Scott, B. E. 1993. The five layers of ambiguity. Byte 18, 174-175.

Simons, G. L. 1984. Introducing Artificial Intelligence. Manchester: NCC Publications.

Sinclair, J. (ed.) 1995.Collins Cobuild English Dictionary 1995. London: HarperCollins Publishers.

The Finnish Formula. Language Industry Monitor 21, 2-6.

Other sources

The homepages of Kielikone Oy at www.kielikone.fi.

APPENDIX

ELÄIMET

Naaraat karkailevat haaremista

Pitkään on uskottu, että poikimisrannoilleen keräytyvät harmaahyljenaaraat parittelevat lähimmän hallitsevan koiraan kanssa. Brittitutkijoiden havainnot eräällä Skotlannin rannikon saarella kumoavat tämän käsityksen, kertoo Science.

Harmaahyljenaaraat nousevat syksyisin vajaaksi kolmeksi viikoksi rannalle synnyttämään yhden poikasen, imettämään ja parittelemaan. Useimmat naaraat palaavat samalle rannalle seuraavina vuosina. Bill Amos kollegoineen sai DNA-analyyseillä selville, että vaikka rannalla pullistelevat voimakkaimmat koiraat keräävät ympärilleen haaremin, useimmat haaremin naaraista livahtavat muualle lempimään.

Lisäksi naaraat näyttävät olevan uskollisia aremmille sulhoilleen. Testatuista 120:stä pennusta vain 29 oli mahtiurosten jälkeläisiä. 120 pennusta löytyi 21 sisarusparia, jotka kulkuriurokset olivat siittäneet samoille naaraille peräkkäisinä vuosina. Sen sijaan hallitsevat urokset siittivät samana aikana vain kaksi sisarusparia.

Naaraan uskollisuus yhdelle kumppanille saattaa vähentää koiraiden aggressioita ja auttaa poikasia pysymään hengissä. Moni pentu nimittäin jyrätään alle, kun koiraat taistelevat haaremista.

ANIMALS

The females keep running away from the harem

It has been believed for a long time that the female grey seals which gather to their breeding shores copulate with the nearest dominant male. The British researchers' observations on an island on the coast of Scotland disprove this idea, Science reports.

The female grey seals get to be less than three weeks to the shore in autumn to give birth to one offspring, to suckle and to copulate. Most females will return to the same shore during the following years. Bill Amos with its colleagues found out with the **DNA** analyses that even though the strongest males that puff themselves up on the shore gather

the harem around them, most of the females of the harem slip elsewhere to make love.

Furthermore, the females seem to be faithful to its shyer fiances. Only 29 of 120 tested cubs were descendants of the powerful males. Out of 120 cubs there were 21 pairs of siblings which the wandering males had begotten during consecutive years by the same females. Instead the dominant males begot only two pairs of siblings during the same time.

The faithfulness of the female to one partner may reduce the males' aggressions and may help the young to stay alive. Many cubs will namely be steamrollered when the males fight for the harem.