

Machine Translation of Various Text Genres¹

Machine translation (MT) has been both praised and criticized since the 1930's when it was first introduced. Today, MT – much improved since then, is a vital tool for the human translator, although not without its problems. One important issue which to our knowledge has not yet been investigated is the success of MT for different text types. In the present study, we compare the performance of German-English machine translation in four different text genres which vary in their structures, using Systran Systems. Systran Company, one of the oldest and most reputable MT producers (dating back to 1968), has been involved with top governmental agencies, such as the US National Air Intelligence Center and the US Air Force's Foreign Technology Division. The texts are analysed with respect to two types of linguistic errors; errors which impede correct transfer of meaning (such as mistranslation of idioms) and errors which merely affect the flow and readability of the texts (e.g., mistranslation of prepositions). These error types can be roughly equated to the traditional measures of intelligibility and fidelity, respectively. Our results show that MT is still limited in its ability to process certain text types, namely those with complex sentence structures, high amounts of pragmatic information and broad semantic domains. In addition, MT tends to produce a number of linguistic errors, most notably the mistranslation of polysemous items. In the final part of the paper, we identify the most frequent linguistic errors and the texts genres MT is best suited for. The theoretical implications of the methodology proposed and the hypotheses investigated constitute the core of the contribution made by this paper.

Key words: German-English translation, machine translation, text genres, translation, Systran Systems.

1. Introduction and aims

What can be said about machine translation (MT) at the present time? “The resulting literary style [from machine translation] would be atrocious and fuller of ‘howlers’ and false values than the worst that any human translator produces” said J.E. Holmström in his famous report for Unesco (1951). However, the idea that machines could be employed in some form of translation survives today and has arguably done so for the last 60 years since its birth in the mid 1930s, which incidentally predates that of the computer (Hutchins, unpublished, p. 1).

This study contributes to the field of MT by taking the novel approach of comparing various text genres with respect to different linguistic errors found when these are translated using machines. Will different text types have different error types or different proportions of the same error? Are there particular texts that machine translation handles better than others? Should we expect to receive “better” machine translations when translating technical texts – whose (semantic) domains are more restrained – than a text such as a short story extract? While the notion that MT may yield better results in certain text genres than in others is not entirely new, to the best of our knowledge no other study investigates precisely this issue.

¹ The preliminary version of this paper was presented at the 7th Language and Society Conference of the New Zealand Linguistic Society, in November 2002, Hamilton, New Zealand.

The core strength of this paper lies with its methodological considerations and with the hypotheses put forward throughout the research. The complete validation of the results presented in the latter part of the work necessitates the support of other similar studies (which employ this methodology) on a much larger scale, with longer texts, possibly a larger number of languages and better quality translation software.

Our goal is not that of directly comparing the quality of human translation versus machine translation, but rather that of obtaining a compilation of frequent linguistic errors which occur in MT. This is of critical importance for future improvements to MT systems and also as a comprehensive guide for optimal usage of this translation tool. According to Bennett (2003, p. 157), while “linguistics has not solved the problems of MT, [...] it can help the researcher to reach solutions, by offering a range of observations, techniques and theories that may be adopted and extended within the MT enterprise”.

In order to achieve these goals and answer the questions above, we compare different text genres with respect to sentence structure, amount of pragmatic information contained and semantic scope. Section 3.1 explains the text genres in more detail and the basis on which they were chosen.

We then identify the kinds of linguistic errors that one might expect in the raw output obtained from fully automated machine translation (FAMT) and classify them in two groups: vital errors which impede the accurate translation of meaning and errors which merely slow the flow and readability of the text, without actually changing or breaking down the intended meaning. The two groups of errors roughly correspond to the traditional notions of *intelligibility* and *fluency*, respectively (White, 2003, p. 216). The specific errors discussed in this paper will be outlined in section 3.3.

The translation of four texts from German into English was done using the SYSTRAN engine currently available on the internet. The types of texts are: technical set of instructions, newspaper extract, popular non-fiction piece and a short story extract.

2. Background, history and attitudes

The mid 30s saw the birth of a revolutionary idea, supported by the French-Armenian Georges Artsrouni and the Russian Petr Troyanskii: machines used as translation tools (Hutchins, unpublished, p. 1). The latter not only envisioned an “automated bilingual dictionary” (ibid), but also a “scheme for coding interlingual grammatical roles” (ibid). To imagine that all this happened before the invention of the first computer is almost unthinkable.

In spite of all this, the first impact of these ideas became a reality in 1946 when the mathematician Warren Weaver developed them further (Hedden, 2000 – unpublished, p. 2). Today, we find various kinds of machine translations: machine-assisted human translation (MAHT), computer-assisted translation (CAT), human-assisted machine translation (HAMT) and fully automatic machine translation (FAMT). They all have different uses. Some are used exclusively as aids to human translators, while others are a quick way of obtaining a translation (admittedly of lower quality) at ‘the touch of a button’ (Hedden, 2000 – unpublished, p. 1-2). It is the latter kind of machine translation – FAMT – that we are concerned with in this paper.

Attitudes towards machine translation have developed in both directions: positive and negative. It is of vital importance that reasons for the two sides are presented and that MT is understood in the context of what it is designed for rather than as a mere competitor of human translators.

2.1 Negative points of view

The report that ALPAC (Automatic Language Processing Advisory Committee) produced in 1966 came close to killing MT altogether. It claimed that “there is no immediate or predictable prospect of useful machine translation” (Hutchins, unpublished, p. 1). In 1975, Chomsky himself wrote “as for machine translation and related enterprises, they seemed to me pointless as well as probably quite hopeless” (Chomsky, 1975, p. 40 in Gross, 1992, p. 109).

Gross (1992, p. 109-110) explains that there are two main problems with machines and natural languages. The first is related to context and cultural issues. Computers are (currently?) not able to perceive the contextual and pragmatic information that humans can. Similarly, they are not aware of cultural differences which often surface in linguistic exchanges. Arnold *et al.*, (1994, p. 108) also support this point.

The second is to do with the function of language. Conveying meaning is just one use of human language, but in addition there are several others such as: humour, establishing solidarity, sharing emotions and feelings without needing to convey any actual information, etc. Plays, poetry, advertising, newspaper articles and songs lyrics are difficult to translate even for humans, hence for the reasons mentioned above, computers cannot provide quality translations for any of these pieces, according to Gross (1992, p. 111).

Ambiguity, idioms, collocations and structural and lexical differences between the source language (SL) and the target language (TL) are also valid concerns in MT (Arnold *et al.*, 1994, p. 105). These problems are discussed in more detail in section 3.3.

2.2 Positive points of view

Among the multitude of complaints MT still receives some well-deserved praise. As Crystal (1987, p. 350) puts it: “The pendulum has begun to swing back again in recent years, following the major intellectual and technological developments of the 1970s in linguistics and computing”.

“A computer’s inability to acquire, comprehend and rationally apply real-world knowledge in this way does not render MT useless as a production tool” says Newton (1992, p. 4). Even though it cannot be used “indiscriminately” (Newton, 1994, p. 7), according to Newton, MT can be of benefit because machines are consistent in interpretation and terminology, do not skip paragraphs or pages and do not make some of the wrong assumptions that even experienced human translators might make.

Gross (1992, p. 109-101) agrees with Newton and adds that machines are faster, more cost effective and more accurate than their human counterparts. He argues that given a particular subject domain, computers are competitive if limited to that domain in the same way that human translators specialised in a particular translating domain can be made efficient. He also warns that the apprehensiveness towards MT may be a side effect of people’s general fear of machines and technology, rather than a valid criticism to the actual field itself (1992, p. 116).

The question that arises is the following: if MT is seen as being successful at all, then what proof is there? The only proof that can be brought comes from current projects where MT is used. The English-Spanish translating system used by the Pan American Health Organization (PAHO) is indeed one such example. It is a machine translation engine, built especially to deal with medical data. This system is the primary mode of translation employed by the company (Vasconcellos and Bostad 1992, p. 64). PAHO undertook a study in order to evaluate the quality of MT versus human translations which showed MT to be more effective (Vasconcellos and Bostad, 1992, p. 67).

Another system which uses MT tools is the Air Force's Foreign Technology Division (FTD) in America. A partially edited text which was translated from Russian into English was assessed by independent sources. Their conclusions about the comparison between the machine translated parts vis-à-vis the human translations were the following: "While the [human] translation read somewhat more smoothly, it seemed to use inappropriate or erroneous terminology more often than the [machine] translation did. Consequently, we relied primarily on the [machine] translation, using the [human] translation mainly for reference." (Vasconcellos and Bostad, 1992, p. 76).

According to Somers (1992, p. 194), by far the most success enjoyed by MT is related to its use in the French weather predictions translations. "... the METEO system, which translates daily more than 30,000 words of weather bulletins from English into French at a cost of less than 0.5¢ (Canadian) per word, with an accuracy rate of 95 per cent" performs a "boring" job which could have difficulties finding human translators willing to do it, explains Somers (1992, p. 194).

A further application of MT is also that of web-searches and translation of web pages. As the internet is flooded with new users from various backgrounds and cultures, linguistic demands are also increasing and languages become barriers to communication. As a consequence, many search engines (Google, AltaVista, Yahoo and others) provide machine translation services to improve (if not eliminate) this problem.

2.3 Evaluation of MT

The diverging negative attitudes which continue to cast a shadow over the field of machine translation can in part be attributed to the difficulty in objectively and uniformly evaluating the quality of MT. So difficult is this task, that a whole new field seems to have emerged, one that has allegedly grown bigger than that of MT itself (cf. Wilks, 1994).

MT evaluation has proven "hard" because no so-called "gold-standard" exists. Creating it would involve investigating a series of different parameters, many of which cannot be easily obtained or objectively measured: quality of translation, cost effectiveness, speed, user-friendly-ness, reliability of the system and the list goes on. The present study is concerned exclusively with quality of translation and hence we restrict the discussion to this parameter. The difficulty in obtaining a reliable and universal measure of MT quality comes from two major factors:

- (1) For any given text, there is not one, unique "right" translation (even human translations of the same texts differ to each other which makes it hard to create a benchmark).
- (2) Rating the quality of the output is not clear-cut. The nature of language (malleable, continuous, creative) and its functions (establishing solidarity, expressing emotions,

creating distance, demanding information, to name a few) do not fit tightly in a mathematical model.

These factors are compounded by various other minor factors, such as different users having different needs from the MT system (for some “rough” translations are enough, for others precise translations are required).

According to White (2003, p. 222), there are six major “models” of evaluation types: feasibility tests, internal evaluation, declarative evaluation, usability evaluation, operational evaluation and comparison evaluation. A brief description of each evaluation type is given below (see White, 2003, for a more detailed discussion).

Feasibility measures test whether a given theory or approach can aid and improve automated translation. Internal evaluations check that the components of a particular MT system adequately perform the translation task. They can be of two kinds: black-box (analysing the quality of the source text against the target text, without being concerned with what actually goes on “inside” the MT system itself) or glass-box (concerned with the performance of the individual components of the MT system). The results of declarative evaluations consist in measures of quality of translations. These are obtained by counting the number of errors found in the target text, or alternatively by comparing the target text to a human translation of the source text, or by simply getting an intuitive measure of the quality of the translation (as a percentage or score). The ALPAC report (1966) was based on this kind of method. Usability methods can be understood as aiming to answer the question “can the MT system be used adequately by the users it was designed for?”. Operational methods address the issue of whether the system is worth using at all – the JEIDA report is one such example (Nomura and Isahara, 1992). Finally, as the name suggests, comparison methods compare various MT systems to each other.

Evaluation methods need not fall into one single model type, they can also combine various aspects of the models described above to give rise to new methods. For instance, project DARPA (started in the mid-1990s and still in use today) combines black-box techniques with the declarative method (native speakers of the target language are asked to rate the output on a five-point scale in terms of the following parameters: adequacy, fluency and informativeness), to obtain an overall measure of quality of three MT systems, each translating across different language pairs (White, 2003, p. 235-238).

Here, we use a black-box method for comparing the performance of the Systran Systems across four different text genres. The advantage in choosing this method consists in its portability, the evaluation test is not system-specific and it can hence be adapted and used with other translation software. This technique is combined with the declarative method, for the purpose of obtaining an analysis of the linguistic errors which occur in the output.

Once an evaluation method is adopted, the next step is that of choosing smaller, specific tests which act like building blocks and contribute meaningful bits of information to the more general question. Each test – termed *metric* – is analysed at the level of expressions or sentences in the output and receives a score in terms of numerical or Boolean values. Collections of metrics can be used to give an idea of the *attributes* (White, 2003, 216) of a particular MT system. For instance, the notions of *intelligibility* and *fidelity* (or adequacy and fluency) are commonly used attributes in MT evaluation.

According to the ISLE classification system (2001) intelligibility measures reflect how “fluent or understandable” a given translation may be, while fidelity measures are indicative of the “accurateness and completeness of the information conveyed”. However, the individual metrics used to obtain these measures vary considerably since not everyone agrees on what it means for a target text to exhibit adequate intelligibility and fidelity to its source. It has also been claimed that alone, measures of intelligibility and fidelity do not provide enough information (see Correa, 2003). According to Reeder, ratings of specific intuitive measures of quality (such as fidelity and intelligibility) “did little to inform either users or developers about the linguistic abilities of the system” and “the search continues for meaningful metrics which correlate with an overall score of success” (2001, p.1). The set of metrics proposed in this study aims to investigate hypothesised linguistic weaknesses of the Systran System and to provide an overall comparative measure of quality among the various text genres, as will be discussed in section 3.3.

2.4 Direction of study

Stepping aside from the various positive and negative attitudes to MT, we find a completely different point of view, taken by those who deny any competition between human and machines in the field of translation. This is not the purpose of MT, they claim. Hutchins states that “there is no doubt that computer-based translation systems are not rivals to human translators, but they are aids to enable them to increase productivity in technical translation or they provide means of translating material which no human translator has ever attempted” (Hutchins, 2001, p. 5).

One innovative aspect of the present research is bourn out this approach to the problem of MT. In extension to Hutchins’s view, we propose that once established as a tool, MT should be assessed and improved just like any other instrument (such as an automated airplane pilot or an automatic car). In accordance with this vision, our research attempts to make suggestions about possible and needed improvements for MT, from a linguistic point of view², as well as to comment on the types of text genres that MT is best suited for. To the best of our knowledge, little (if any) work specifically focuses on the comparison of various machine translated text genres.

3 Methodology

Several factors related to the design of this study will be addressed in this section. The following three sub-sections are dedicated to explaining how the texts were chosen, what translation tools have been used and which set of (linguistic) metrics were investigated.

3.1 Sample of texts

The number of different text genres analysed was limited to four. They ranged from the most “pragmatic” (containing an abundance of contextual information) to the least context bound. On the one hand, we wanted to choose a kind of text that will have little pragmatic information, relatively

² The artificial intelligence part and computer programming areas are left to the experts. Here we only discuss linguistic concerns.

concise and even short (where possible) sentences and a narrow semantic scope. This is why we chose a set of technical instructions. Naturally, at the other end of the scale, it was desired to have a highly pragmatic, stylistically and semantically rich, elaborate text, i.e., an extract from a classic short story.

Before considering the remaining two text genres, it is worth noting that the technical set of instructions was not written or subjected to a controlled language (CL) style. A controlled language is an “explicitly defined restriction of a natural language that specifies constraints on the lexicon, grammar, and style” (Nyberg *et al.*, 2003, p 245)³. The main purpose of CLs is that of creating a “one-to-one correspondence between word forms and concepts” throughout a given text, and hence increasing “readability and comprehensibility” for both humans (in human-oriented CL) and/or machines (in machine-oriented CL) cf. Nyberg *et al.*, (2003, p 246-248). Indeed, it turns out that the application of CLs does seem to improve the quality of both human and machine translated texts (*ibid*). In this study, the major focus of investigation is the success rate of MT in various text genres, not its performance with CLs. For this reason, the technical set of instructions was left in its original form.

Between the two extremes of the technical set of instructions and the short story, a newspaper article was also chosen, since newspapers thrive on context-bound information (it is presumed that the readers already know of previous developments of events around the world). However, newspapers use a lot of short, effective sentences. The last item chosen was a piece from a popular women’s magazine. This is a more neutral kind of text, as far our criteria is concerned. Pragmatic information creeps in occasionally, but is not always there. Sentences are neither elaborate, nor short. They are usually a good mix of both.

Table 1 summarizes the desired properties of the chosen texts.

Text genres	Sentence types	Amount of pragmatic information	Semantic domain/scope
Technical set of instructions	Short	Little	Very limited
Popular magazine extract	Combination of long and short	Neutral	Neutral
Newspaper article extract	Many short and effective	Lots	Very broad
Short story extract	Long, elaborate	In abundance	Very broad

Table 1. Text genres

³ For more information on CLs see Huijsen (1998) and Nynberg *et al.*, (2003), and for a discussion of the application of CLs specifically to the field of machine translation see Knops and Depoortere (1998) and Mitamura (1999).

The length of the texts translated was limited to circa 200 words. As we wanted all the texts to finish at the end of a sentence, some texts may go over the 200 words boundary. The limit was decided upon for the following reasons:

- First, it was desired to have a decent length which would give us a good range of linguistic errors, or at least the most frequent ones.
- Secondly, as the aim of the paper is not that of comparison between different styles of translations, but instead it is one of investigating the optimal contexts for MT use and the most urgently needed improvements to this method of translation; we are not interested in obtaining an evaluation measure for MT. Hence, there is no need to perform statistical analyses on the texts and the word size can be limited to a manageable one, as long as the previous condition is satisfied.

3.2 Translating tools

As discussed in section 1, the texts were translated from German into English. The machine translation engine used is SYSTRAN systems. According to Crystal, “one of the best-known MT systems, SYSTRAN, was developed in the US with particular reference to Russian–English translation – for example it was used to translate Russian into English during the Apollo–Soyuz space project” (1987, p. 350). Since then, SYSTRAN was extended to other languages, German being one of them. The engine used is part of a free service and can be found at the following internet address: <http://www.systranet.com/systran/net>. This service is a well established, reputable MT engine, which a private consumer could use if in need of a quick translation at no cost.

The texts were submitted in a simple text format and the translation was instantaneous. The translations obtained were discussed with a human translator with native knowledge of the German language.

3.3 Set of metrics

The linguistic errors under investigation in this study were chosen in accordance with the typologies suggested by Catford (1965) and Arnold *et al.*, (1994).

Both pieces discuss the issue of ambiguity. This type of problem may show itself in various forms. We concentrate here on two of them, namely lexical gaps and polysemy. We continue to use the terminology proposed by Arnold *et al.*, i.e., lexical gaps. However, we will differentiate between *gaps with superfluous words* (exemplified in 3.3.1), where a word in SL (German) does not require a word in TL (English), but is nevertheless wrongly included by the machine translation, and *gaps which require an entire phrase, not just a word* (as in 3.3.2), where a word in SL requires an entire phrase in TL, but the machine fails to fulfil this.

3.3.1	SL	<i>Ich</i>	<i>muß</i>	<i>gehen</i>	<i>nach (PREP.)</i>	<i>Hause.</i>
	TL	‘I	must	go	-	home.’
3.3.2	SL	<i>Ich</i>	<i>kann</i>	-	<i>Deutsch.</i>	
	TL	‘I	can	speak	German.’	

Similarly, we also distinguish between two types of polysemous items (words in SL which have two or more meanings in TL). In the first category are those words which have not been translated by the appropriate variant, but in spite of this, the overall meaning is preserved from contextual information. This type of error will be referred to as *polysemy with preserved meaning*. For instance, the sentence *Das kochende Wasser ist nicht bereit* can be translated as ‘The **cooking** water is not ready’, which despite being intelligible is not entirely correct. A better translation would be ‘The **boiling** water is not ready.’ The second category contains those words whose meaning has been mistranslated and consequently, the general meaning is thereby completely lost (i.e., the German word *Papiere* can be translated into English as ‘securities’ or ‘newspapers’ or simply ‘papers’). These will be included in the category of *polysemy with loss of meaning*.

Arnold *et al.*, (1994) also mention the problems of idioms and collocations in MT. We will follow their terminology and use the term idiom to refer to specific phrases whose meaning cannot be guessed from that of the individual words representing them, such as the German saying *eine Flasche zu sein* (lit. ‘to be a bottle’) which means ‘to be a loser’. In contrast, collocations are understood to be units whose meaning can be guessed from the individual parts, but “what is not predictable is the particular words that are used” (Arnold *et al.*, 1994, p. 119), as for example the phrase *to have a break* is translated into German as ‘eine Pause zu machen’ (lit. *make a break*), not ‘eine Pause zu haben’ (to have a break).

The other linguistic problems investigated are: complete loss of translation (words are left in the SL with no attempt of translation) and structural differences between the two languages (breaching of syntactic and word order rules).

After performing the translations, it was observed that a specific linguistic error type which had only received little attention in the literature, namely the mistranslation of prepositions, appeared consistently across all text genres. For example, the German preposition *von* was translated as ‘of’, instead of the required ‘by’ (*von Ulrich* means ‘by Ulrich’, not ‘of Ulrich’). It was hence decided to include it in the analysis as a separate category (see section 5.2 for a more detailed discussion of this error type).

Table 2 summarizes the linguistic problems discussed in decreasing order, from the ones which contribute *most* to the loss of meaning; to the ones that have the least impact on it. The items that appear on the same line are considered to present an equal amount of damage to the conservation of meaning.

	(1) No translation given	
	(2) Idioms	(2) Polysemy with loss of meaning
	(3) Gaps requiring an entire phrase, not just a word	
	(4) Structural difference between languages	
	(5) Collocations	(5) Polysemy with preserved meaning
	(6) Prepositions	
	(7) Gaps with superfluous words	
Most damaging to meaning conservation		
Least damaging to meaning conservation		

Table 2. Sample of linguistic errors

Each error was recorded so that the raw number of errors is obtained (as given by Table 3, section 4). If exactly the same word was translated incorrectly, hence appearing as an error, more than once within the same text, then that error was only recorded once in the table constructed for that particular extract.

Two main groups of errors were identified in Table 2: those which impede the accurate translation of meaning – these can be roughly equated to an overall intelligibility score; and those which merely impede the flow, readability or completeness of information conveyed – this second group of errors have to do with fidelity measures to the original SL texts. Errors classified as types 1, 2 and 3 (in the table) belong in the first group and errors 4, 5, 6, and 7 are in the second. Hence gaps with superfluous words would not be considered to affect the correct transfer of meaning from the source language to the target language, whereas idioms are thought to have a strong effect on it.

4. Results

The results (exact number of error occurrences) obtained for the three text genres are presented in Table 3. Perhaps not surprisingly, the fourth text genre (the short story extract) was the least adequately translated text genre. The text was so poorly translated that no coherent and meaningful results could be obtained from its translation: no single sentence in the TL made sense, as will be exemplified in section 4.4. Each text type is discussed individually in further detail in sections 4.1-4.4.

Linguistic errors	Occurrences (number of words)		
	Technical Set of instructions	Newspaper extract	Popular non-fiction extract
No translation given	4	5	4
Idioms	0	0	1
Polysemy with loss of meaning	0	6	12
Gaps requiring an entire phrase, not just a word	1	3	2
Structural differences between SL and TL	7	6	14
Collocations	0	0	6
Polysemy with preserved meaning	14	13	16
Prepositions	2	4	5
Extraneous words which require a gap	1	1	8
Totals	29	38	68

Table 3. Results obtained for the three text genres

4.1 Technical set of instructions

The technical set of instructions chosen are instructions for the use of a coffee machine (produced by Bialetti). This text type had the least number of linguistic errors. As shown in Table 3, there were no idioms or collocations and no polysemous words were mistranslated to the point that meaning was actually lost.

By far the most frequent error was that of polysemous words being translated outside the correct context. We give three instances of this in 4.1.1:

4.1.1

<u>SL</u>	<i>Fügen Sie dem Kaffee keine Extrakte, Kakao oder sonstigen Zusätze bei...</i>
<u>Translation given</u>	‘Do not attach to the coffee excerpts , coca or other additions ...’
<u>Correct translation desired</u>	‘Do not add to the coffee extracts , cocoa or other additives ...’

The pairs of words given in 4.1.1 are close to being synonyms and it is the context (i.e., that of preparing coffee) which selects the most appropriate lexical item. In spite of the occasional “wrong” choices made by the machine, the context is also the one that saves the situation by suggesting the intended meaning to the reader.

Prepositions were also a problem. Examples 4.1.2 and 4.2.3 list the two that were mistranslated.

4.1.2

<u>SL</u>	... Wasser <i>muss langsam aber mühelos</i> durch den Kaffee fließen könnten
<u>Translation given</u>	‘...water must be able to flow slowly however easy by the coffee’
<u>Correct translation desired</u>	‘...water must be able to flow slowly however easily through the coffee’

4.1.3

<u>SL</u>	... <i>nehmen Sie die Kaffeemaschine vom Herd</i>
<u>Translation given</u>	‘...take the coffee machine of the stove’
<u>Correct translation desired</u>	‘...take the coffee machine off the stove’

The structure differences between SL and TL were as following:

- Two German separable verbs (verbs composed of two parts: the main verb and a preposition which appears at the end of the sentence) are treated as if they were prototypical verbs,
- One adverb is placed in the wrong place in the sentence (i.e., word order not respected in TL),
- Four verbs are placed in the wrong place in the sentence (word order not being respected in TL).

The words which were left in their original form, i.e., no attempt was made to translate them into English were: *trichterförmigen* ‘funnel-shaped’, *gemahlenem* ‘ground-up’, *Oberfläche* ‘face (surface)’ and *fest* ‘fixed’. Something peculiar happened to the last word in the list. It was not only left un-translated, but it was in fact left out altogether from the English translation.

4.2 Newspaper extract

The newspaper piece is an extract from an article taken from Die Zeit which discusses the present political situation in Afghanistan. It is the text with the second best error score, after the technical set of instructions. Here, we also find no problems related to idioms or collocations.

The most frequent error is the same as in the technical set of instructions: polysemy without loss of meaning. However, there is an increased number of polysemy with loss of meaning and we give some examples of this below:

4.2.1

SL *Wie viele Leichen mussten verscharrt werden unter der Bezeichnung ...*
Translation given 'Like many corpses had to be verscharrt under the designation....'
Correct translation desired 'How many corpses had to be buried under the designation....'

4.2.2

SL *...spielte sie mit ihren Freunden im Hof ihres Mietshauses, ...*
Translation given '...played it with its friends in the yard of their dwelling house,....'
Correct translation desired '...she played with her friends in the yard of their rented house, ...'

Another type of linguistic error which severely affects meaning is that of words in SL requiring a whole phrase in TL, as opposed to a single word. Here is an example from this text type:

4.2.3

SL *...es kann ausgehöhlt werden durch die Art der Kriegführung und durch die Opferzahlen*
Translation given '...it can be scooped out by the type of warfare and by the victim numbers'
Correct translation desired '...it can be uncovered by the manner in which the war is conducted and by the victim numbers'

Some of the prepositions mistranslated are given in examples 4.2.4 and 4.2.5.

4.2.4

SL *von Ulrich Ladurner*
Translation given 'of Ulrich Ladurner'
Correct translation desired 'by Ulrich Ladurner'

4.2.5

SL *Am 17. Oktober*
Translation given 'To 17. Oktober...'
Correct translation desired 'On the 17th of October...'

The structural differences between SL and TL in this piece are:

- One adjective is placed in the wrong place in the sentence (word order not being respected in TL),
- One genitive construction from SL is left out in the TL,
- Verbs (three) are placed in the wrong place in the sentence (word order not being respected in TL),
- One adverb is placed in the wrong place in the sentence (word order not being respected in TL).

The words which were not translated at all were the following: *Passanten* ‘passers-by’, *marktschreiern* ‘market screams’, *Menschengewimmel* ‘swarming crowds’, *verscharrt* ‘buried’ and *Oktober* ‘October’.

4.3 Popular non-fiction piece

This text is an extract of an article from a German women’s magazine, Brigitte, discussing health and fitness tips for busy, career-oriented women. Its translation was worse than all the other texts, excluding the short story extract.

The most frequent error was also polysemy with preserved meaning, just as in all the other pieces translated. The next most common problem was that of structural differences between the two languages. Below is a list of the types of errors found:

- Eight German separable verbs (verbs composed of two parts: the main verb and a preposition which appears at the end of the sentence) are treated as if they were prototypical verbs,
- One adverb is placed in the wrong place in the sentence (i.e., word order not respected in TL),
- Four verbs are placed inappropriately in the sentence (word order not being respected in TL),
- One adjective is placed in the wrong place in the sentence (word order not being respected in TL).

Similarly, there are a high number of polysemous items with loss of meaning. Some of the most significant examples are shown below:

4.3.1

SL <Gut fürs Wohlbefinden und die Gesundheit>, meint Hella Thomas.
Translation given ‘<Well for the well-being and the health>, Hella Thomas **means.**’
Correct translation desired ‘<It is good for well-being and health>, **thinks** Hella Thomas.’

4.3.2

SL ...übermittelt dem Gehirn ständig die **entsprechenden** Befehle ...
Translation given ‘...constantly transmits to the brain the **appropriate** instruction’
Correct translation desired ‘...constantly transmits to the brain the **corresponding** instruction...’

4.3.3

SL ...hat man Appetit auf Obst oder Gemüse (Vitamine!) und sehnt sich nach **Ruhe**.

Translation given ‘...one has appetite on fruit or vegetable (vitamine!) and longs itself for **peace**’

Correct translation desired ‘...one has appetite for fruit and vegetables (vitamins!) and longs for **rest**’

This translation contains one idiom and seven collocation problems. It is the only text genre which produced these types of linguistic errors. Examples 4.3.4 – 4.3.6 give some of these.

4.3.4

SL Ist eine Erkältung im **Anmarsch**...

Translation given ‘If a cold is **in the advance**...’

Correct translation desired ‘If a cold is **on the way**...’

4.3.5

SL **Im Job** ist ein Spaziergang nach dem Mittagessen eben nicht immer **drin**.

Translation given ‘**In the job** a walk is not after the lunch evenly always **in it**.’

Correct translation desired ‘**At work**, a walk at lunchtime is not always **on**.’

4.3.6

SL **Gut** fürs Wohlbefinden und die Gesundheit

Translation given ‘**Well** for the well-being and the health’

Correct translation desired ‘**It is good** for well-being and health’

Similar to the other texts, prepositions also create problems in this translation (we also saw one example in 4.3.3 already, where the output contained ‘on’ instead of ‘for’):

4.3.7

SL ...Training oder Bewusstseinsübungen nach der Feldenkrais-Methode
schärfen die Wahrnehmung **für** die eigenen Bedürfnisse...

Translation given ‘...training or consciousness exercises, according to the Feldenkrais method
sharpen the perception **for** the own needs...’

Correct translation desired ‘...training or consciousness exercises, according to the Feldenkrais
method sharpen the perception **of** one’s needs...’

The German words which were not translated were: *morst* ‘to use the morse code’, *Vitamine* ‘vitamins’, *Schokoriegel* ‘candy bar’, *Energiebringer* ‘energy enhancers’ and *hecheln* ‘to pant’.

4.4 Short story extract

The translation of this text – a short story called *Die Flucht*, written by Reiner Maria Rilke – was indeed virtually unreadable. There was not one single coherent sentence in the translation. The

disappointment started with the title itself: ‘The flight’ instead of ‘The escape’. The following is a representative sample of what the machine translation looks like:

‘By the multicolored *glasfenster* [‘glass window’] over the main altar follow the evening jet, broadly and simply, like the old masters the faded colours of the level carpet represent, into the main ship and refurbished it on the *verkuendigung Mariens* [‘the annunciation of Virgin Mary’].’

There is one remark to be made about this piece, in spite of its lack of intelligibility. The word “color” from the first line is spelled “colour” in the following one. It appears that there are different varieties of English used within the same translated text. This is an easily fixable error and one which raises some doubts about the quality of the translation engine.

Due to the low quality of the translation, there is no point in even counting up the errors made.

5. Discussion of results

This section discusses the main findings in the paper: first, with respect to the text genres chosen, and then, in reference to the specific linguistic errors analysed.

5.1 The text genres

All translations contained at least one or more of the linguistic errors investigated.

The short story extract was impossible to translate. There was virtually nothing to be gained from using MT to translate it, as there was not one single sentence translated correctly. It can be assumed however that no MT system would ever be designed for the purpose of translating this type of text genre. The technical set of instructions, on the other hand, scored best in terms of linguistic errors made (i.e., had the least number of mistakes). The newspaper extract followed and the popular non-fiction extract came third overall.

If we were to turn a blind eye to the number of occurrences of polysemy without loss of meaning, we estimate that the technical set of instructions would be translated adequately enough for someone to understand and be able to follow the instructions given for the use of the coffee machine. The newspaper article on the other hand would pose additional problems with the six occurrences of polysemy with loss of meaning. Finally, the translation of the popular non-fiction extract is very limited and may not provide the reader with much more than the mere gist of the topic discussed in the text. These observations seem to suggest that in cases where a translation of lower quality is sufficient (i.e., for the purpose of getting “basic” information quickly, finding out the topic/theme of a piece of text, or deciding whether the text is of interest to the reader before investing more time and funds into translating it), all three text types can be translated using automated translating tools – with varying degrees of success (technical instructions would yield more intelligible and fluent translations than say a newspaper article). However, when a higher quality of translation is needed, it is only texts containing technical information or sets of instructions which give adequate (or close to adequate) results.

The results presented in this study agree with the hypothesis presented by Gross (1992, p. 109-101), who claims that texts with “limited subject domains” can be translated successfully by machines. These findings show that texts genres which contain little pragmatic information, clear and short sentences and restricted semantic domains achieve best results in MT. In contrast, those which are heavily dependent on contextual information, have long, elaborate sentences and a broad semantic domain are poorly translated by machines.

A back-translation was performed for the best translated text (the set of technical instructions), whereby the first translation obtained (German → English) was re-translated into the original language (German). The newly obtained translation was almost better than the first one⁴ (German → English). There were no words left in the SL (this time English) and there were less linguistic errors identified overall (under 10). This finding may not be at all surprising since the vocabulary and constructions in the translated English text were generated from the computer programme itself – hence it might be expected that it would recognize and translate them accordingly. On the other hand, this also shows that the linguistic errors identified in the translated English (German → English) version did not constitute a barrier in the following task of re-translation (English → German). Another observation which can be made from the back-translation is the fact that translating from English into German (as opposed to from German into English) may prove a more fruitful task, as the MT system appeared to have produced a output of better quality in that direction. In other words, the language pairs chosen and the direction of translation may turn out to yield better or worse translations. This hypothesis is left to be investigated by future research.

The area of back-machine-translation is sometimes used as an evaluation for checking the quality of the “first” (or original) translation. This is particularly common in cases where the team evaluating the MT system does not have experts on both the SL and the TL (and hence cannot evaluate the quality of the first translation directly). However, the results of this study suggest that one cannot make direct assumptions about the reliability of a particular MT system solely based on the method of back-translation, as this type of test may hide certain flaws of the system (such as lexical gaps in the TL, as we saw earlier).

5.2 The linguistic errors

All linguistic errors investigated appeared in at least one or more of the texts analysed. The most frequent error for all text genres is that of polysemy with preserved meaning. This is not a problem which threatens the accurate transmission of meaning from one language to another. It merely slows the readability and flow of the piece.

Other frequent errors are (in the order of frequency, from the most frequent to the least frequent) structural differences between SL and TL and polysemy with loss of meaning.

⁴ There were however problems regarding the difference in symbols from English and German alphabets (the ‘ä’, ‘ö’, ‘ü’ and ‘ß’ were not identified correctly in my version of Word – Microsoft Office 2000).

Separable verbs contribute uniformly, across text genres, to the category of structural differences between languages. This seems to be a problem which the machine is able to deal with to a certain extent, as there are cases where separable verbs were correctly identified and hence accurately translated, as shown in 5.2.1. This type of error applies specifically to German → English translations (and other languages which allow separable verb constructions).

5.2.1	<u>SL</u>	<u>TL</u>
	setzen ...ein	insert
	fallen ...auf	are noticeable
	hörten...auf	stopped

Word order problems such as the correct placement of adverbs and adjectives were also present, even if more infrequently. They did not affect meaning as much as the mistranslation of separable verbs because all the needed components of the sentence are present, whereas in a separable verb construction, the meaning of the verb “unit” is different than that of the original “single” verb.

The number of words which were not translated at all stayed relatively constant (3 words – 5 words per piece) for each text type. Lexical gaps requiring an entire phrase (not just a word) were also present at a constant rate (1 word – 3 words per piece) in all text genres. These two errors were infrequent and did not pose problems to the preservation of meaning. The most infrequent linguistic errors found were idioms and collocations. They were only present in the popular non-fiction text. This can be explained by the fact that the popular non-fiction text was the only one to contain these types of constructions.

Among those errors which contribute to the loss of meaning, the most frequent error was polysemy with loss of meaning across all text types, with the exception of the set of technical instructions (where it was the error of “no translation given” – words left in the original German form).

The category of mistranslated prepositions contributed to the list of linguistic errors found. It was present in *all* text genres. This problem has not gone completely unnoticed. Studies by Japkowicz and Wiebe (1991) and Decrozant and Voss (1999) note that locative prepositions pose great problems for an MT system translating from English into French. The main thesis developed by Japkowicz and Wiebe (1991) is that locative prepositions are not conceptualised in English as they are in French, hence causing their mistranslation. As a possible solution, they propose the use of “knowledge representations of conceptualisation of objects” based on models outlined by Lakoff (1987) and Herskovits (1986) and claim that this could be potentially extended to languages other than French and English. Nübel (1996) observed the difficulty in translating prepositions within spoken discourse (also using MT systems) and proposed that both “lexical semantic knowledge” and “context information provided by a dialogue model” are needed in order to “capture the context dependence of preposition meanings” (p 10).

As suggested by Nübel (1996), prepositions pose a lot of difficulty, especially for MT systems because of their being heavily dependent on contextual information. In a sense, one could argue that prepositions belong to the category of collocations because they occur in phrases whose meaning can be guessed from the individual parts representing them, but the process of choosing the parts (i.e., the prepositions) is still a mystery, e.g., in English it is said that a book is written ‘**by** Jules Verne’,

whereas in German the book is ‘**from** Jules Verne’ (*von Jules Verne*). As attested by the previous example itself, our findings show that locative prepositions are not the only type of prepositions which pose problems to MT. There are also relational prepositions, attributive prepositions, benefactive prepositions and so on, all of which constitute a problem for MT systems. Due to the high number of occurrences, there is a need for linguistic as well as computational strategies which can eliminate this type of error in the future.

There were also potential linguistic problems which the machine translation avoided. For example, negations were translated appropriately and modal verbs and auxiliaries appeared in the correct form and the appropriate place in the translations given. However, no comprehensive survey of these was obtained since, as Reeder suggests in her 2001 paper, “negative exemplars are very indicative of MT improvement needs” and “focusing on the positive aspects of language, they [previous studies done in this fashion] miss the real value in analysing the errors generated by the systems”.

6. Concluding remarks and future directions

The present study constitutes a novel approach to analysing machine translation quality across various text genres, with respect to different linguistic errors. The findings show that machines produce better translations of technical sets of instructions than of other types of texts. Translations of highly contextual pieces containing complex sentences and broad semantic domains, such as short story extracts, are virtually unintelligible. There is little gain from submitting pragmatic and linguistically complex material to a machine and comparing its output to a human translation. This is not the purpose of an MT system (at least not for now).

Our findings suggest that so long as MT systems are used with texts which have restricted semantic scopes, relatively little pragmatic information and simple sentence structure, the translations obtained, though unable to retain the fluency of the original texts, do an adequate enough job of preserving the original content of the SL. This means that for users who want to get “the gist” of a given text or for those who are content to obtain fast but “imperfect” translations, MT may prove an effective, valuable and not to mention cheap tool. For instance, MT could enable people to search the internet and buy products sold overseas, by companies which may not trade in their language; or it could give direct access to resources and services (booking a hotel room in Paris without speaking French) to people who do not belong to the same linguistic community and have no/little knowledge of each other’s language.

Some of the types of linguistic errors which need urgent consideration are polysemous items which lead to loss of meaning, structural differences such as inseparable verbs (German), mistranslation of prepositions and last but not least, words which are not translated at all in the target language. An issue which remains for future research is how different machine translation methods compare with respect to the linguistic errors investigated here. Are there perhaps MT methods which cope better with prepositions or do they all fall short of translating them appropriately?

Machines can be valuable tools in translation and they perform better in some text genres than in others. However, there remain particular linguistic problems which need to be solved before

progress in this field can be made. Ultimately, it is the kinds of texts which machines are used to translate that determine whether the output is at all useful and meaningful and as with many other computer-related tools (such as the internet) it is up to the user to make the most gains from it. Such gains come from informed use, from being aware of both limitations and strengths of the tool at hand.

Acknowledgements

The author would like to thank W.S. Allan for his suggestions, guidance and help with the compilation of the text genres table (Table 1, section 3), A. Kirkness for his help with the translations, S. M^cNamara for making suggestions regarding the topic and D. Starks for constructive criticism. Thanks are also due to the anonymous referees for their meticulous, patient and helpful comments which helped improve the paper significantly.

References

ALPAC: 1966, *Language and machines: computers in translation and linguistics*, A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington DC.

Arnold, D.J., Balkan, L., Meijer, S., Humphreys, R.L. and Sadler, L.: 1994, *Machine translation: an introductory guide*, London, Blackwells-NCC.

Bassnett-Mcguire, Susan: 1990, *Translation studies*, Great Britain, The Chaucer Press Ltd.

Bennett, Paul: 2003, 'The relevance of linguistics for machine translation', in Somers, Harold (ed), *Computers and translation – a translator's guide*, Amsterdam, Philadelphia, John Benjamins, pp. 143-160.

Catford, J.C.: 1965, *A linguistic theory of translation*, London, Oxford University Press.

Chomsky, Noam: 1975, *The logical structure of linguistic theory*, Chicago, University of Chicago Press.

Correa, Nelson: 2003, 'A fine-grained evaluation framework for machine translation system development', *Proceedings to the MT Evaluation Summit IX: Towards systematizing MT Evaluation*, New Orleans, Louisiana, USA.

Crystal, D.: 1987, *The Cambridge encyclopedia of language*, Cambridge, Cambridge University Press, pp. 350-351.

Decrozzant, Lisa and Voss, Clare: 1999, 'Dual use of linguistic resources: evaluation of MT systems and language learners', *Computer Mediated Language Assessment and Evaluation in Natural Language Processing Symposium*, University of Maryland, College Park, MD.

Dorr, Bonnie J.: 1993, *Machine translation: a view from the lexicon*, Cambridge, The MIT Press.

Gross, Alex: 1992, 'Limitations of computers as translation tools', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 96-130.

Hedden, T.D.: 2000, *Machine translation: a brief introduction*, Website at http://www.he.net/~hedden/intro_mt.html, – unpublished, consulted on March, 20th, 2002.

Huijzen, W. O.: 1998, 'Controlled language – an introduction', *Proceedings of CLAW 1998*, Pittsburgh, pp. 1-15.

Herskovitz, A: 1986, *Language and special cognition: an interdisciplinary study of the prepositions of English*, Cambridge University Press, Cambridge, MA.

Holmström, J.E.: 1951, *Report on the interlingual scientific and technical dictionaries*, Paris, Unesco.

Hutchins, John: 2001, 'Machine translation and human translation: in competition or in complementation?', *International journal of translation*, 13:5-20.

Hutchins, John, *The history of machine translation in a nutshell*, Website at <http://ourworld.compuserve.com/homepages/WJHutchins/Nutshell.htm>, unpublished, consulted on March, 20th, 2002.

ISLE: 2001, *MT Evaluation classification, expanded classification*, Website at <http://www.isi.edu/natural-language/mteval/2b-MT-classification.html>, consulted on April, 24th, 2004.

Japkowicz, Nathalie and Wiebe, Janyce: 1991, 'A system for translating locative prepositions from English into French.', *29th Annual Meeting of the Association for Computational Linguistics*, pp 153-160.

Knops, U and Depoortere, B.: 1998, 'Controlled language and machine translation', *Proceedings of CLAW 1998*, Pittsburgh, PA.

Lakoff, G: 1987, *Women, fire, and dangerous things: what categories reveal about the mind*, University of Chicago Press, Chicago.

Mitamura, T.: 1999, 'Controlled language for multilingual machine translation', *Proceedings of Machine Translation Summit VII*, Singapore, pp 46-52.

Newton, John: 1992, 'Introduction and overview', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge.

Nomura, H and Isahara, J.: 1992, 'The JEIDA report on machine translation' in AMTA, *Proceedings of the AMTA workshop on MT Evaluation: basis for future directions*, San Diego.

Nübel, Rita: 1996, 'Knowledge sources for the disambiguation of prepositions in machine translation', *Proceedings of the Workshop on Future Issues for Multilingual Text Processing*, Pacific Rim International Conference on Artificial Intelligence, Cairns, Australia.

Nyberg, E., Mitamura, T. and Huijsen, W.: 2003, 'Controlled language for authoring and translation', in Somers, Harold (ed), *Computers and translation – a translator's guide*, Amsterdam, Philadelphia, John Benjamins, pp. 245-281.

Paris, Jean: 1992, 'Translation and creation', in Arowsmith, W. and Shattuck, R. (eds), *The craft and context of translation*, Texas, The University of Texas Press, pp. 57-67.

Pugh, Jeanette: 1992, 'The story so far: an evaluation of machine translation in the world today', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 14-32.

Reeder, Florence: 2001, 'In one hundred words or less', *Proceedings to the MT Evaluation Workshop, MT Summit VIII*, Santiago de Compostella, Spain.

Schubert, Klaus: 1992, 'Esperanto as an intermediate language for machine translation', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 78-95.

Somers, Harold L: 1992, 'Current research in machine translation', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 189-207.

Vasconcellos, M and Bostad, D.A.: 1992, 'Machine translation in a high-volume environment', in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 58-77.

White, John. S: 2003, 'How to evaluate machine translation', in Somers, Harold (ed), *Computers and translation – a translator's guide*, Amsterdam, Philadelphia, John Benjamins, pp. 211-244.

Winter, Werner: 1971, 'Impossibilities of translation', in Arowsmith, W. and Shattuck, R (eds), *The craft and context of translation*, Texas, The University of Texas Press, pp. 68-82.

Wilks, Yorick: 1992, 'SYTSRAN: it obviously works but how much can it be improved?' in Newton, John (ed), *Computers in translation – a practical appraisal*, London, Routledge, pp. 166-188.

Wilks, Yorick: 1994, 'Keynote: traditions in the evaluation of MT', in Vasconcellos, M (ed), *MT evaluation: basis for future directions. Proceedings of a workshop sponsored by the National Science Foundation*, San Diego, California.

References for the translated texts

- **Technical instructions piece** the Bialetti coffee machine consumer instructions.
- **Newspaper article extract** Die Zeit, consulted on March 8th 2002, from <http://www.zeit.de/>
- **Popular non-fiction extract** Brigitte, consulted on March 12th 2002, from <http://www.brigitte.de/gesundheit/gesundheit/koerpersignale.html>
- **Short story extract** Rilke, Maria, Reiner. Die Flucht, consulted on March 19th 2002, from http://www.unix-ag.uni-kl.de/~kasperek/Rilke/Die_Flucht.html

References for the translating engine

- Main web page for SYTRAN is at <http://www.systransoft.com/>
- The translating service used can be found at <http://www.systranet.com/systran/net>