

Has machine translation improved? some historical comparisons

John Hutchins

[WJHutchins@compuserve.com]

Abstract

The common assertion that MT systems have improved over the last decades is examined by informal comparisons of translations produced by operational systems in the 1960s, 1970s and 1980s and of translations of the same source texts produced by some currently available commercial and online systems. The scarcity of source and target texts for earlier systems means that the conclusions are consequently tentative and preliminary.

1. Introduction

A question frequently asked by those new to the field is whether machine translation (MT) has improved – in the last five years, or in the last ten years, or in the last twenty, etc. In many respects, the answer is quite easy. It is obvious that programs for automatic translation run much faster than in the past because computers are faster. Equally obvious is that text input is much easier than in the days of punched cards and paper tape, and that the output is much easier to read than in the days of almost illegible computer printouts all in upper case. MT systems are also becoming cheaper by the year, even by the month. Finally, it is plain that more and more large organizations can and do save translation costs by using MT systems as aids in the production of draft translations – which was almost inconceivable twenty years ago.

However, these are not the improvements that are usually meant when the question is asked. The ‘real’ question is whether the *quality* of machine translation has improved. Some will say that it *must* have. Others of an older generation insist that MT has the same problems and throws up the same errors as it did in the 1960s and 1970s. Many, however, will perhaps share my intuition that while progress in quality is not evident during say the last ten years, there are discernible advances since the early 1980s and definite improvements since the ALPAC report of the mid 1960s. But what is the evidence for this impression, strong as it may be. Perhaps we are being misled or dazzled by the genuine and widespread technical advances.

This paper will be an attempt to look at the output of MT systems from earlier periods of the field’s history (before the early 1990s) and to see whether today’s systems would perform any better on the same texts. It is an informal (i.e. unsystematic) effort at historical comparison – as yet, there seems to be no available methodology for longitudinal evaluation. But it is also unsystematic because it is necessarily based on partial evidence. It is surprisingly difficult to find examples of actual ‘raw’ translations by operational MT systems before the last decade – from an extensive search through my substantial archive. Most often, we find fragments (individual sentences) and not extended texts, and where we do find extensive translations, they are usually given without the original source texts – and so no comparisons with other systems (of the time or later) are possible. Many examples in the literature are from research prototypes – sometimes selected in order to give the most favourable impression of the system’s capabilities – these have not been used, since developers would legitimately protest at the use of pre-release versions of their systems. A final restriction is that comparisons with modern systems are limited to those readily available to the non-specialist user – this means commercial systems for personal computers and on-line translation services. No doubt, researchers will claim that more recent developments in corpus-based techniques will produce better results, but as far as the general public is concerned, these potential improvements are not yet real.

Comparisons are made on the basis of ‘raw’ (unedited) output. It is recognised that this should not be the sole criterion for evaluating the usefulness of systems (e.g. for producing publishable quality texts), but it is only by examining unedited output that we can judge whether MT engines themselves have improved – and not post-editing facilities. Whatever the context of use (dissemination, assimilation, etc.) it must surely be true that the better the quality of ‘raw’ translation the more “useful” the system. Among many other factors, evaluations should be able to say whether a new product (or a new

version of an old product) is or is not better in terms of translation quality than an older product (or an older version of the same product.)

The language directions covered are from Russian, French and German into English, because they have the longest histories. The restriction to operational systems means: IBM and Georgetown in the 1960s, Systran in the 1970s (other systems such as GETA-Ariane and SUSY were experimental), Logos, METAL and some PC systems (ALPS, Weidner, PC-Translator, Globalink) in the 1980s and early 1990s. Examples for most have been located (except early ones for many PC systems). For the modern current systems the choice has been limited to 'personal' MT systems (such as Systran Personal, Personal Translator and Promt) and online services (Babelfish, Lycos, Reverso, Promt, FreeTranslation, InterTran), since I have no access to current 'enterprise' systems such as Compendium, Systran Enterprise, m²T (previously Logos), Reverso Corporate, Transcend Corporate, etc. In this paper it is not possible to reproduce extensive extracts. Here will be found only a few illustrative (hopefully representative) examples, with brief comments – and usually only the 'best' versions of modern systems are given. Fuller extracts of the source texts (originals and translations), and additional examples of translations by current systems – and examples of some other language directions – will be found in a much extended form of this paper available on my website (<http://ourworld.compuserve.com/homepages/WJHutchins>).

2. Russian to English

2.1. IBM-USAF system

The first genuine output of a MT system is provided by the Russian-English system developed for the US Air Force Foreign Technology Division (Dayton, Ohio) and operational from 1959. Here is an example (quoted by Masterman and Kay 1960): the original Russian (1) and the English output (2).

(1) [original]: В арсенале средств научно-атеистической пропаганды, материалистического воспитания трудящихся появилось новое оружие – журнал “наука и религия”. Вышел в свет его первый номер. Это сборник содержательных, разнообразных, с интересом читающихся материалов, острие которых направлено против религиозных суеверий и предрассудков.

(2) [IBM-USAF, 1959]: In arsenal means scientific-atheistichesky propaganda, materialistic training worker appeared new weapon – magazine “Science and Religion”.

Issued in light it first number. This collection containedkh, various, with/from interest reading matterevilv, sharp that directed against religious suervery and prejudice.

The translation by the IBM-USAF system is poor and almost unreadable: the system lacked some items of vocabulary, failed to interpret some case endings (-kh), produced an inexplicable translation of *материалов* as *matter* and *evilv*, created some aberrant English forms (*issued*), and in general failed to analyse the Russian syntax. For the same text, Systran and PROMT produce the following:

(3) [Babelfish, 2003]: In the arsenal of the means of scientific atheistic propaganda, materialist training of workers did appear new weapon periodical is “Science and religion”. Its first number appeared. This is the collector/tank of the meaningful, diverse, with the interest read materials, whose point is directed against the religious superstitions and the prejudices.

(4) [PROMT Online, 2003]: In an arsenal of means of scientific - atheistic propagation, materialistic education of working the new weapon - magazine " a science and religion " has appeared. His(its) first number was published. It is the collection substantial, various, with interest of read materials which edge is directed against religious superstitions and prejudices.

Neither Systran nor PROMT are good translations, although the gist can be extracted. Some failings of the IBM-USAF system are still present. Interpretation of the pre-nominal participle construction *с интересом читающихся материалов* (meaning *materials to be read with interest*) still eludes both Systran and PROMT. In some cases Systran is more successful than PROMT (*point* vs. *edge*), in others PROMT is better (*collection* vs. *collector/tank*). But both are clearly better than the IBM-USAF version.

A later version of the IBM-USAF system was demonstrated in 1964 at the New York World's Fair, but the quality was suspiciously flawless (Bowers and Fisk 1965).

2.2. Georgetown system

Apart from the IBM system, the only systems of the first “generation” of MT to come into operation were those delivered in 1963 and 1964 to EURATOM at Ispra in Italy, and the US Atomic Energy Commission at Oak Ridge National Laboratory by the team at Georgetown University. An example from EURATOM (Perschke 1968):

(5) [original]: О возможности возникновения нестабильностей в плазме, захваченной поперечным магнитным полем.

Исследованы условия возникновения и развития нестабильности в плазме, захваченной поперечным магнитным полем и прошедшей через диафрагму.

Показано, что при заземлении металлической диафрагмы нестабильности плазмы не развиваются.

The EURATOM output comes from about 1967:

(6) [EURATOM, 1967]: Concerning the possibility of emergence of instabilities in plasma, which was captured by cross magnetic field.

Studied the conditions of emergence and the developments of instability in plasma, which was captured by cross magnetic field and which passed through diaphragm. Showed, that upon the grounding of metallic diaphragm of instability plasmas are not developed.

Obvious problems are verb-initial constructions (*studied, showed*) and the difficulties of selecting appropriate articles – Russian has none – and the treatment of prepositions (in particular, *of* in genitive constructions). The performances of Systran (7) and PROMT (8) are generally improvements:

(7) [Babelfish, 2003]: On the possibility of the appearance of instabilities in the plasma, seized by transverse magnetic field.

Are investigated the conditions for appearance and development of instability in the plasma, seized by transverse magnetic field and which passed through the diaphragm. It is shown that during grounding of metallic diaphragm the instabilities of plasma are not developed.

(8) [PROMT, 2003]: About an opportunity of occurrence нестабильностей in the plasma seized by a cross magnetic field.

Conditions of occurrence and development of instability in the plasma seized by a cross magnetic field and past through диафрагму are investigated. It is shown, that at grounding metal диафрагмы instability of plasma do not develop.

Systran still occasionally fails to deal with verb-initial constructions (*Исследованы...* translated as *Are investigated...*); both Systran and PROMT translate захваченной as *seized* rather than the better EURATOM rendition (*captured*); PROMT does not recognise the plural *instabilities* (*нестабильностей*) and lacks dictionary entries for *diaphragm*, which probably accounts for the mistranslations in the second paragraph. However, in general, Systran and PROMT are improvements. (Similar comments apply to examples from ORNL – see expanded paper.)

2.3. Systran system

The Russian-English version of Systran was the first of the Systran systems to become operational. It was installed in 1970 at the USAF Foreign Technology Division, replacing therefore the IBM system above – because it gave better results. We should expect, therefore, to find that the quality of Systran output in the 1970s to be high.

One example of Systran output in Bruderer (1978) is suspiciously almost flawless. More typical

of the capability of Systran in the mid 1970s is probably one of the translations from 1976 used by Knowles (1979) in his ‘error analysis’ of Systran for the European Communities.

(9) [original]: Вертолет, летательный аппарат тяжелее воздуха с вертикальными взлетом и посадкой, подъемная сила в котором создается одним или несколькими (чаще двумя) несущими винтами... Вертолет взлетает вертикально вверх без разбега и совершает вертикальную посадку без пробега, неподвижно висит над одним местом, допуская поворот вокруг вертикальной оси в любую сторону, производит полет в любом направлении со скоростями от нуля до максимальной.

(10) [Systran, 1976]: A helicopter, a flight vehicle heavier than air with vertical by takeoff and landing, lift in which is created one or by several (more frequent than two) rotors. ... A helicopter takes off upward vertically without a takeoff and it accomplishes vertical fitting without a path, motionlessly ‘will hang’ above one place, allowing rotation around a vertical axis to any side, flight in any direction at speeds is produced from zero to the maximum.

In the first sentence, a copula *is* should have been inserted – not, however, an easy problem when translating from Russian, where often a dash or comma indicates equation. The presence of *by* should have ideally been eliminated, and *lift in which* would be more comprehensible as *where the lift...* More serious are the translations in the second sentence of посадку as *fitting* (instead of *landing*) and of без разбега as *without a path*. The repeated *fitting* in the third sentence might still be confusing, but in general the translation serves well as a rough version.

How then does the current Systran system compare?

(11) [Babelfish, 2003]: Helicopter, aerodyne with the vertical by the takeoff and landing, lift in which is created by one or several (more frequent than two) rotors... Helicopter takes off vertically upward without the takeoff/run-up and accomplishes vertical fitting/landing without the path/range, motionlessly it will hang over one place, allowing rotation around the vertical axis to any side, flight in any direction with speeds of from zero to maximum is produced.

The use of the uncommon *aerodyne* in place of the earlier *flight vehicle* is a surprise. As in 1976, the first sentence lacks a copula *is*, it retains the intrusive *by* and it still has *lift in which*. There are however some improvements in the second sentence where alternative translations are offered: *run-up* as well as *take-off* and *landing* as well as *fitting* (in this case correct). A step backward, however, is the placing of *is produced* at the end of this sentence.

Does its rival PROMT perform better?:

(12) [PROMT, 2003]: The helicopter, the flying device is heavier than air with vertical rise and landing(planting),

elevating force in which is created by one or several (is more often to two) bearing(carrying) screws ... the Helicopter flies up vertically upwards without start and makes vertical landing(planting) without run, motionlessly hangs above one place, supposing turn around of a vertical axis in any party(side), makes flight in any direction with speeds from zero up to maximal.

In the first sentence, *flying device* is better than *aerodyne* (but not than the earlier (10) *flight vehicle*), and the copula is present, although misplaced and giving a confusing structure. The use of *screws* instead of *rotors* is an unhappy choice; likewise, of *rise* instead of *takeoff*, and of *without start* instead of *without take-off*. On the other hand *landing without run* in (12) is better than the puzzling *landing without path* (retained essentially in the current Systran (11)). But, in the same sentence, PROMT is worse with the output of *supposing turn around of a vertical axis* instead of *allowing rotation around vertical axis* as in both (10) and (11).

In general, the current Systran is marginally better than the 1976 Systran version, but PROMT is perhaps not. It seems that we cannot point to the substantial improvements in Russian-to-English MT in the last 25 years that Knowles (1979) had expected would come.

3. French to English

In 1976 the European Communities acquired Systran for translating its growing volume of documents. One of the first systems was for French-to-English translation. Beaven (1998) has a number of examples illustrating definite progress of the system over a ten-year period.

3.1. Systran system

Here we look at an example from the early 1980s given by Wagner (1985) [for others see the expanded paper]:

(13) [original] Où en est-on, environ un an après la réunion du Comité permanent de l'emploi? A l'époque, les organisations syndicales avaient estimé que la Commission européenne prenait le problème par le petit bout de la lorgnette, sans le situer dans le contexte d'une stratégie globale et d'une politique industrielle plus dynamique. Il faut partager le travail, réduire le temps de travail, mais ce n'est là qu'un palliatif.

(14) [Systran 1983] How far was, approximately a year after the meeting of the Standing Committee on Employment? At the time, the trade-union organizations had felt that the Commission had taken the problem by the small end of the lorgnette, without locating it in the context of a total strategy and of a politique industrial more dynamic. It was necessary to share work, to reduce the working time, but it was there only one palliative.

Apart from one lexicon omission (*politique*), the system at this time could not deal with the common French expression *où en est...*; nor with the complexities of the less common idiom *ce n'est là que...* The current Systran version (15) is an improvement on the whole – the main exception being a mistranslation of the last sentence due to a failure to treat the comma before *réduire* as a sign of coordination. The treatment of *où en est...* is still clumsy, but it is now intelligible; however, the phrase *ce n'est là que...* is still a problem. It is, of course, not surprising that it can give no English equivalent for the metaphor *prendre... par le petit bout de la lorgnette* – now (15) or in the past (14).

(15) [Babelfish, Lycos, Systran Personal 4.0] Where is one, approximately a year after the meeting of the Standing Committee on Employment? At the time, the trade-union organizations had estimated that the European Commission took the problem by the small end of the spyglass, without locating it in the context of a total strategy and a more dynamic industrial policy. Work should be shared, to reduce the working time, but it is only one palliative there.

The output from Reverso On-line (64) is also better on the whole than the 1983 Systran (14). There is idiomatic rendition of *où en est...* although spoiled by the intrusive *on* before *approximately*. Like the current Systran systems it fails with the metaphor and with the phrase *ce n'est là que...*

(16) [Reverso, 2003] Where are we, on approximately one year after the meeting of the permanent Committee of the employment(use)? In the time, unions had considered that the European Commission took the problem by the small end of the opera glasses, without placing him(it) in the context of a global strategy and a more dynamic industrial politics(policy). It is necessary to share the work, to reduce the working time, but it is there only a palliative.

However, the online MT system Free Translation is markedly inferior – a less acceptable version for *où en est...*, a mistranslation of *à l'époque*, and of *une politique industrielle plus dynamique*, as well as (just like Systran and Reverso) a failure with *ce n'est là que*:

(17) [FreeTranslation, 2003] Where some is one, about a year after the meeting of the permanent Committee of the employment? To the era, the union organizations had considered that the European Commission took the problem by the small end of the lorgnette, without situating it in the context of a global strategy and of a political industrial one more dynamic. It is necessary to divide the work, reduce the work time, but this is not there that a palliative one.

But another online service InterTran produces a virtually incomprehensible result, with inexplicable insertions (1994 and *thanksgiving*):

(18) [InterTran, 2003] Where at am on , about a an after her sitting any Board continuous about the application? HAS the era , the organizations trade unionism had

appraised as her Commission Europe took him issue by the snip any leer at , without him locate in 1994 context d'une strategy aggregate and d'une political industrial anymore dynamic. He must share him employment , curtail time about employment , but thankssgiving n'est there than it is to a palliatif.

3.2. Systran system: another example

Another example from Wagner (1985) illustrates other problem areas:

(19) [original] M. le Président rappelle que le problème de la réduction du temps de travail a été étudié à la réunion de Munich. Différentes thèses s'affrontent: pour les syndicats, la réduction du temps de travail contribuera à supprimer le chômage, mais les employeurs pensent qu'elle supprimera des emplois soit en renchérissant les coûts, soit en accroissant la productivité. Il serait souhaitable de poursuivre aujourd'hui cette discussion en laissant de côté tous les présupposés idéologiques. Pour commencer, il convient de demander au représentant de la Commission européenne, qui a réussi à rejoindre Luxembourg malgré un mouvement de grève à Bruxelles, s'il souhaite compléter l'exposé qu'il avait présenté à Munich.

In 1983, the EC version of Systran produced:

(20) [Systran 1983] The Chairman reminded the meeting that the problem of the reduction of the working time had been studied to the meeting of Munich. Various theses faced themselves: for the trade unions, the reduction of the working time would contribute to suppress unemployment, but the employers thought that she would suppress an employment either by increasing the costs, or by increasing the productivity. It would be desirable to continue today this discussion by leaving some on the side of all ideological presupposed them. To start, it was advisable to require from the European Commission representative, which had succeeded in joining Luxembourg despite a strike movement in Brussels, if it wished to supplement the statement that it had presented in Munich.

There is an awkward translation of *s'affrontent*, an understandable mistranslation of *elle* as *she*, a failure to treat the idioms *laisser de côté* and *il convient de*, and the (less tolerable) mistranslations of *qui* as *which* and *il* as *it* (twice).

Current Systran systems produce a better rendition of *s'affronter*, manage to correctly translate *elle*, but still fail with *qui* and *il* in the final sentence. They also introduce an unnecessary *to* before *ask*. Nevertheless, there is overall a definite improvement:

(21) [Babelfish, Lycos, Systran Personal, 2003] Mr. the President points out that the problem of the reduction of the working time was studied with the meeting of Munich. Various theses clash: for the trade unions, the reduction of the working time will contribute to remove unemployment, but the employers think that it will remove employment either by increasing the costs, or by increasing the productivity. It would be desirable to continue this discussion today by leaving side all the presupposed ideological ones. To start, it is advisable to ask to the European Commission representative, which succeeded in joining Luxembourg in spite of a movement of strike in

Brussels, if it wishes to supplement the talk that it had presented in Munich.

On the other hand, Reverso (22) and FreeTranslation (23) are hardly better than Systran in 1983. FreeTranslation has the same mistakes as in the older Systran (20), it also fails to recognise *productivité* and *idéologiques* and it has a more awkward rendition of *il convient de*.

(22)[FreeTranslation, 2003] Mr. the President recalls that the problem of the reduction of the work time was studied at the meeting of Munich. Different theses confront themselves: for the unions, the reduction of the work time will contribute to eliminate the unemployment, but the employers think that she will eliminate from employments be while increasing the costs, be while increasing the productivité. It would be desirable to follow today this discussion while leaving out all them presupposed idéologiques. To begin, it suits to request the representative of the European Commission, that succeeded rejoining Luxembourg despite a strike movement to Brussels, if it wishes to complete the exposition that it had presented to Munich.

The version from Reverso is worse in some parts than either the 1983 Systran (20) or the current Systran and FreeTranslation outputs (21, 22): *reminds* instead of *recalls* or *points out*, *syndicates* instead of *unions*, *kill(abolish)* instead of *remove* or *eliminate*. Nevertheless the overall result is not significantly worse.

(23) [Reverso, 2003] Mr. President reminds that the problem of the reduced working week was studied in the meeting of Munich. Various theses are in confrontation: for syndicates, the reduced working week will contribute to kill(abolish) the unemployment, but the employers think that it will kill(abolish) employments(uses) either by making more expensive(by adding) the costs, or by increasing the productivity. It would be desirable to pursue this discussion today by leaving aside all the ideological presuppositions. To begin, it is advisable to ask to the representative of the European Commission, who managed to join Luxemburg in spite of a strike action in Brussels, if he wishes to complete the statement which he had presented in Munich.

Overall, we can conclude that French-to-English translation by the Systran, FreeTranslation and Reverso systems are all improvements on the translations of the early 1980s by Systran. However, there are systems on the market today which are no better and, on occasions, far worse – e.g. by InterTran (not shown here) – than their predecessors twenty years ago.

4. German to English

Although research on German began during the first decades of MT activity (e.g. at Massachusetts Institute of Technology and the University of Texas), operational systems for the language did

not appear until the 1980s, and non-corporate commercial systems not until the early 1990s. Some of the earliest examples of German-to-English MT are therefore from pre-prototype systems, e.g. the METAL and Systran systems in the 1970s (Bruderer 1978: 260-269 and 298-299).

4.1. Logos system

However, we do have output from the Logos system installed at SAP in the early 1980s. An example from 1983 (Tschira 1985) is:

(24) [original] Im Geschäftsleben werden die Geschäftsvorfälle regelmässig durch Belege dokumentiert, wie etwa Buchungsbelege, Aufträge, Bestellungen. Die in ihnen enthaltenen Daten werden – in Gegensatz zu den Stammdaten – als “Bewegungsdaten” bezeichnet.

In herkömmlichen EDV-Anwendungs-Systemen wurden oft die Daten eines Geschäftsvorfalles getrennt bearbeitet und gespeichert. In den SAP-Systemen vermeiden wir diese unnatürliche Aufteilung und speichern einen Geschäftsvorfall statt dessen einheitlich als sog. “Dokument”, um dem Sachbearbeiter jederzeit einen Überblick über den gesamten Vorfall zu ermöglichen.

(25) [Logos 1983] In the business-life, the business transactions are documented regularly by document/ vouchers, as for example posting documents, orders, orders. The data included in them are designated as movement data – in contrast to the master data -.

The data of a business transaction were often processed and stored separately in conventional EDP-application-S-systems. In the SAP-systems, we avoid this unnatural breakdown and store a business transaction instead of which than sucked uniform. At any time, to make possible document, about the person in charge an overview of the entire occurrence.

It may be noticed in the first paragraph that Logos could not distinguish between *Aufträge* and *Bestellungen*, and that it misplaced the embedded phrase. In the second paragraph, Logos fails to recognise the genitive link -s-, the standard phrase *statt dessen*, and, most seriously of all, it fails to recognise *sog.* as an abbreviation of *sogenannt (so-called)* with a consequential nonsensical translation *sucked* and mis-parsing of the whole sentence.

However, current Systran systems are no better (26) at dealing with *sog.* in the final sentence, and the third sentence is also distorted.

(26) [Babelfish, Lycos, Systran Personal 4.0] In the business life the business transactions are documented regularly by vouchers, as for instance vouchers, orders, orders. The data contained in them become - in contrast to master data - when "designates transaction data".

In conventional EDP application systems the data of a business transaction were often worked on and stored separately. In the SAP systems we avoid this unnatural allocation and store a business transaction instead of its uniformly as if sucked. "document", in order to at any time make for the specialist possible an overview of the entire incident.

Reverso (27) succeeds with *sog. (sogennant)*, but fails elsewhere: it does not recognise *EDV-Anwendung* and, like Logos (25), fails to make the link to *systems*. More seriously, *getrennt* is interpreted as a full verb and not as a past participle adverb (*separately*), and the translations of *Vorfall* as *incident* rather than *transaction* would probably be misleading. However, there is correct recognition of *statt dessen* as adverbial *instead*.

(27) [Reverso On-line] In the business life the commercial incidents are documented regularly by vouchers, approximately like vouchers, orders, orders. The data contained in them are called - in contrast to the master data - "movement data".

In conventional EDV Anwendungen to systems the data of a commercial incident were treated often separated and were stored. In the SAP systems we avoid this unnatural subdivision and store a commercial incident instead uniformly as so-called. "Document" to make possible a view about the whole incident for the specialist at any time.

FreeTranslation version (28) also succeeds with *sog.*, but again there are misleading interpretations of *Vorfall*, *getrennt*, and *statt dessen*, and in addition a mistaken translation of *Stammdaten*.

(28) [FreeTranslation] In the business life, the business occurrences are documented regularly through records, as well as about entry record, orders, orders. The data contained in them are designated – in contrast to the tribe data – as “movement data”.

In conventional EDP-uses-systems, the data of a business occurrence were separated worked and stored often. In the SAP systems, we avoid this unnatural distribution and store a business occurrence instead of its uniformly as so-called. “Document” in order to enable the expert always an overview over the total occurrence.

With Personal Translator (29), the success with *sog.* is outweighed by a number of other errors, of which the most obvious are the morphologically odd *finishedly* and *storedly*:

(29) [Personal Translator PT] In the business life the business transactions are regularly documented by pieces of evidence such as booking pieces of evidence, orders, orders. The data contained in them are described as "movement data" in contrast for the master data.

The data of a business transaction were finishedly and storedly often differentiated in conventional EDP application systems. In the SAP systems we avoid this unnatural division and store a business transaction instead of this one uniformly as a so-called "document" to make a summary of the complete occurrence possible for the clerk any time.

Overall, we may conclude that there has been only slight progress since 1983.

4.2. Globalink system

Finally there is the question whether there is any discernible improvements in the last decade with systems designed for ‘home use’, i.e. primarily for

those who simply want to find out the gist of a text (letter, form, etc.). In 1990, an MT evaluation group at Essex University made an assessment of the translation quality and operational performance of the commercial Globalink GTS system (University of Essex 1991). An extract from an official letter:

(30) [original] Wir haben Ihren Antrag auf einem Platz für den oben genannten Kurs bearbeitet und ich freue mich, Ihnen mitteilen zu dürfen, dass Sie für das Studienkolleg-Programm des am 1. Oktober 1990 beginnenden Studienjahres zugelassen sind. Sind Ihre Leistungen während des Studienkollegs zufriedenstellend, können Sie im Oktober 1991 ein dreijähriges Studium mit dem Abschluss "Bachelor's degree" im Bereich der Wirtschaftswissenschaft beginnen.

In 1990, Globalink (31) produced almost complete nonsense for the first sentence:

(31) [Globalink GTS, 1990] We have processed Your proposition on a place for the above named course and I please me, You communicate to be allowed to, that You are admitted for the @@Studienkolleg – program the at the 1. October 1990 beginning study year. If Your performances Are during the study course of lectures satisfactory, can begin You in the October 1991 a three-year study with the conclusion "Bachelor's @@degree" in the area of the economy sciences.

The current Systran versions (32) are definite improvements, despite the inexplicable and surprising *are pleased I*, the inappropriate *certified* (instead of *admitted*) and the misplaced *beginning* in the first sentence.

(32) [Babelfish, Systran Personal 4.0, 2002] We worked on your request on a place for the course specified above and are pleased I to be allowed to communicate to you that you are certified for the preparatory course program on 1 October 1990 of the academic year beginning. If your achievements are satisfying during the preparatory course, you can begin a three-year study in October 1991 with the conclusion "Bachelor's degree" within the field of the economic science.

FreeTranslation and Reverso outputs (33, 34) are also both better than the 1990 Globalink. Like Systran, the first sentences have confused structures, but, on the other hand, unlike Systran, both succeed in correctly interpreting the pre-nominal participle construction *des am 1. Oktober 1990 beginnenden Studienjahres*, placing *beginning* immediately after *studies year*. Reverso also produces the correct *admitted* (where Systran has *certified*, and FreeTranslation *allowed*). However, FreeTranslation fails to recognise the *if*-clause in the second sentence, which is a major failure. But while Reverso (like Systran) gets this right, it manages to produce an adverb (*satisfyingly*) instead of an adjective (*satisfying*, or better *satisfactory*):

(33) [FreeTranslation] we worked your proposition on a place for the course named above and please me I to be allowed to announce to you that you are allowed for that

studies lecture program of the studies year beginning on the 1 October 1990. Your achievements are during the studies lecture satisfactorily, you can begin in October 1991 a three-year study with the termination "Bachelor's it degree" in the area of the economy science.

(34) [Reverso] We have treated your application(motion) for a place(square) for the course named at the top and I am glad to be allowed to inform you that you are admitted for the study lecture-program of the academic year beginning on the 1st October, 1990. If your performances (achievements) are during the study lecture satisfyingly, you can begin in October, 1991 a three-year-old study with the conclusion " Bachelor's degree " in the area of the economic science.

While Systran, FreeTranslation and Reverso are generally improvements on the Globalink output of the early 1990s, the outputs from Personal Translator PT and InterTran (not shown here) are definitely not.

5. Concluding comments

Historical comparisons are complicated by two major factors: (a) the absence of examples of translations (with their source texts) for systems; and (b) the variable quality of current systems, not just between systems but also from one text type to another within the same system. To a large degree, individuals will come to different conclusions about whether there have or have not been improvements in quality, depending largely on which texts they test and what criteria they use. Nevertheless, it does seem that overall there has been definite progress since the mid 1960s and very probably since the early 1970s. What is more uncertain is whether and where there have been improvements since the early 1980s. The general quality of (unedited) Russian-to-English translation seems to have scarcely improved in the last two decades. For French-to-English there has certainly been a gradual increase in translation quality since the mid 1980s, but perhaps not markedly so in the last decade. For German, however, it does seem that there have been gradual if slight improvements since the mid 1980s, although the overall quality is in general behind that found in translations from French.

The impediments to the improvement of translation quality are the same now that they have been from the beginning: failures of disambiguation; incorrect selection of target language words; problems with anaphora: pronouns (*it* vs. *she/he*), definite articles (e.g. when translating from Russian and French); inappropriate retention of source language structures (e.g. verb-initial constructions (from

Russian), verb-final placements (from German), non-English pre-nominal participle constructions (e.g. *with interest to be read materials* from both Russian and German); problems of coordination; numerous and varied difficulties with prepositions; and in general always problems with any multi-clause sentence.

Here is not the place to suggest how these problems might be solved, or even how the worse cases might be avoided, but it does seem strange that after nearly 50 years of MT research some commercial systems still produce incorrect morphological forms (e.g. *finishedly*), incorrect noun-verb agreements, incorrect adjective-noun orders, and incorrect placements of verbs at the beginnings or at the ends of sentences and clauses. Such apparently easily resolvable errors must undoubtedly contribute to the dissatisfaction of many users of current 'personal' and online MT systems. These are errors which, it would seem plausible to suggest, could be reduced by some form of statistical 'language model', e.g. to produce the right prepositions, to insert articles, to rearrange word orders, and to correct the morphology (cf. Knight and Chander 1994)

Poor quality software is not, of course, unique to MT – while there is a market for low grade products, there will be manufacturers to sell them. For more informed purchasers there is a need for readily available and easily understood benchmarks and consumer evaluations; and the need is more urgent as the years pass. However, this paper has not been concerned with current poor-quality systems but with the question whether there are now commercial and/or online systems producing better 'raw' output than older operational systems. Some tentative answers have been given, but more investigation is desirable.

This paper (and its expanded version) has been a preliminary and informal attempt to discover whether there is any basis for claiming (or implying) that real progress in output quality has been achieved in recent years. The scarcity of published (or otherwise recorded) translation examples from older systems makes the task difficult. But if the MT community is to have any idea of how much (or how little) it has progressed, and if it is to have good records of what has been achieved, it needs to preserve and archive systems and programs of older MT systems, both in order to evaluate their performance as translating systems, and in order that later generations can investigate

how the programs actually worked. (In some cases, this may mean preserving the technology as well – computers, operating systems, disk drives, magnetic tape readers, line printers, paper tape equipment, etc.) As MT has become more commercial, the need to preserve older systems has become more pressing. One message from this paper is, therefore, that steps should be taken now, before it is too late, to preserve all superseded MT software in an archive for future researchers.

6. References and sources of examples

- Beaven, J. (1998): 'MT: 10 years of development', *Terminologie et Traduction* (1998), 1: 242-256.
- Bowers, D.M. and Fisk, M.B. (1965): 'The World's Fair machine translator', *Computer Design*, April 1965, 16-29.
- Bruderer, H.E. (1978): *Handbuch der maschinellen und maschinenunterstützten Sprachübersetzung: automatische Übersetzung natürlicher Sprachen und mehrsprachige Terminologiedatenbanken*. München: Vlg. Dokumentation Saur, 1978.
- Knight, K. and Chander, I. (1994): 'Automated post-editing of documents', *Proc. International Conference on Artificial Intelligence (AAAI)*, 779-784.
- Knowles, F. (1979): 'Error analysis of Systran output – a suggested criterion for the 'internal' evaluation of translation quality and a possible corrective for system design'. In: Snell, B. (ed.) *Translating and the Computer: proceedings of a seminar, London, 14th November 1978* (Amsterdam: North-Holland), 109-133.
- Masterman, M. and Kay, M. (1960): *Mechanical pidgin translation*. Cambridge: Cambridge Language Research Unit. (ML 133).
- Perschke, S. (1968): 'Machine translation – the second phase of development', *Endeavour* 27, 97-100.
- Tschira, K.E. (1985): 'Looking back at a year of German-English MT with Logos'. In: Lawson, V. (ed.) *Tools of the trade: Translating and the computer 5... Proceedings of a conference... 10-11 November 1983*, The London Press Centre (London: Aslib, 1985), 215-235.
- University of Essex (1991): *Assessing a PC-based commercial MT system*. Essex MT Evaluation Group. Colchester: University of Essex Department of Language and Linguistics, January 30, 1991. (Working Papers in Language Processing, no. 19)
- Wagner, E. (1985): 'Rapid post-editing of Systran'. In: Lawson, V. (ed.) *Tools for the trade: Translating and the computer 5. Proceedings of a conference... 10-11 November 1983*, The London Press Centre (London: Aslib, 1985), 199-213.