DOCUMENT RESUME

ED 413 745 FL 024 766

AUTHOR Krauwer, Steven
TITLE Machine Translation: State of the Art, Trends and User Perspective.
PUB DATE 1995-00-00
NOTE 8p.; In: Language Resources for Language Technology: Proceedings of the TELRI (Trans-European Language Resources Infrastructure) European Seminar (1st, Tihany, Hungary, September 15-16, 1995); see FL 024 759.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computational Linguistics; Computer Software; *Discourse Analysis; Foreign Countries; Language Research; *Languages; Linguistic Theory; *Machine Translation; Research Methodology; Research Needs; Structural Analysis (Linguistics); *Translation; Trend Analysis; Users (Information)
IDENTIFIERS Europe (East); *Language Corpora

ABSTRACT

Currently, no machine translation (MT) system is capable of successfully imitating the behavior of a human translator, and there exists no formal description of what an MT system is supposed to do. The biggest problem in practice is disambiguation. However, various types of existing systems do help reduce language barriers, even if they are poor imitations of the human translator. Users of such systems determine the utility of the systems. Three groups of users are distinguished: big companies and institutions; professional and occasional translators; and monolingual users. Big MT systems are becoming more cost-effective. Smaller systems providing high-quality output for restricted tasks and domains and multilingual systems based on spoken dialogue are being developed. Whereas market developments show a trend toward more specialized and restricted systems, the trend in research is to widen the scope of machine translation, moving from sentence to text, focusing on disambiguation based on domain restrictions, and using non-traditional statistical methods. Work is needed to provide lexicological and terminological resources for East European languages, and to create large collections of real-life data, including monolingual and bilingual text corpora. (MSE)

Machine Translation: State of the Art, Trends and the User Perspective

Steven Krauwer

Research Institute for Language and Speech (OTS)
Utrecht University
Trans 10
3512 JK Utrecht
The Netherlands
Tel. +31 30 253 6050
Fax: +31 30 253 6000
E-mail: steven.krauwer@let.ruu.nl

1. Machine Translation, does it exist?

The only honest answer to this question is NO. If we define a Machine Translation system as a system capable of successfully imitating the behaviour of a human translator, we still have a long way to go. As a matter of fact, we do not even know what successfully imitating means: there exists no formal or even semi-formal description of what an MT system is supposed to do. But even in the absence of such a formal description, we know from direct observation that no computer program is capable of delivering translations even remotely approaching the quality of translations made by professional human translators.

The biggest problem in practice is the disambiguation problem. Human beings have no problems in picking out word senses and assigning the proper interpretations and translations to phrases, sentences, and texts, whereas even the most sophisticated computer system would have a hard time identifying the proper referent of the word "they" in the following two sentences (the examples are borrowed from Jerry Hobbs):

(1) The policemen fired at the demonstrating students because they wanted revolution.

(2) The policemen fired at the demonstrating students because they feared revolution.

The reason why we can easily interpret both sentences correctly is that we have more than just our linguistic knowledge to rely on, and it is only when one tries to analyze texts on the basis of linguistic rules only that it becomes clear that more often than not extralinguistic knowledge is needed to assign the appropriate analysis. The real problem is of course not that we could not store somewhere the information that students are more likely to want revolution than policemen and that policemen are supposed to stop revolutions from happening, but rather that it is totally unpredictable what knowledge about our culture, our behaviour, our technology, etc. will be needed in order to disambiguate texts to be translated.

But although it is very tempting to stick to the "NO" answer to the question about the existence of MT and just give up the enterprise (at least for the time being), it somehow feels unsatisfactory to go for this option. After all, the idea that a MT system should simulate the behaviour of a human translator is based on a rather arbitrary interpretation of what MT could be.

If we focus on the underlying problem (language barriers) rather than the instrument traditionally used to solve it (translators) and define an MT system as a system capable of reducing the negative effects of these language barriers, we can identify various types of existing systems which do help us to reduce

these problems, even if they are very poor imitations of the human translator or do not even pretend to be an imitation (and the well-known comparison between birds and aeroplanes comes to mind immediately).

2. Users of MT systems

If we assume that the purpose of MT is to reduce the problems caused by the fact that different people speak different languages, we have to accept that "reduction" means different things to different people. For some people, reduction may refer to the cost of translation; for others, speed may play a more important role, and yet others may want to gain better access to information made available in foreign languages.

Therefore, we have to distinguish a number of classes of users of MT systems. For the purpose of this paper, a distinction into three classes will be made, but it should be kept in mind that more fine-grained classifications are possible.

2.1 Big companies and institutions

Many big companies and institutions operating on an international scale are faced with a huge translation load which justifies an in-house translation service. In such an environment, MT may become an attractive option if at least one of the following criteria is met:

- the translation quality is increased,
- turn-around times are improved, or
- the translation cost per unit is reduced.

If we look at what MT currently has to offer, we get the following picture on the negative side:

- the quality of full MT system output is generally poor, i.e., (post-editing is necessary);
- introduction of MT requires a high initial investment in training of people, acquisition of hardware and software, and in the customisation of dictionaries;
- it requires a different organisation of the work flow;
- the cost of computational support for such systems is high in comparison with e.g., translators using word processors on PCs.

On the positive side we have:

- high speed;
- relatively low exploitation cost in comparison with human translators.

If the volume is big enough, the speed increase or the cost reduction may be significant. Just to mention an example: the system Logos was used by Lexi-Tech to translate 500 000 pages of user documentation for 12 Canadian patrol frigates from English into French. If we assume (optimistically) that a human translator can manage some 8 pages of text per day, this job will require an effort of more than 300 man years. If the price per translated page is estimated at 50 ecu, the overall translation cost will be in the order of 25 Mecu if done by human translators. It is clear that under such circumstances the use of MT may mean the difference between practically and financially feasible and unfeasible. Even if an MT system produces raw translations which require significant human post-editing, an prospective overall cost reduction of, say, 20% will justify a considerable initial investment.

Our conclusion is that this approach will work well for large volumes. If the volume is not large enough, there is little chance that there will be any significant return on the investment.

Typical examples of translation systems used by big companies or organisations:

- Systran, used by the European Commission;
- Metal, used by the Union Bank of Switzerland to translate information technology and telecommunications texts from German into English, for internal use;
- Kielikone, used by Nokia to translate telecommunications customers documentation from Finnish into English.

2.2 Professional and occasional translators

Professional translators are in a different position, since translation is their core business and their specific area of expertise. They often work on a free-lance basis or in small companies. Their annual turnover does not usually justify major investments, and they do not normally want their jobs to be taken over completely by their computers. Occasional translators have even less reason to invest in expensive MT systems.

The main requirements of professional and occasional translators are:

- increase of productivity,
- better quality, and
- low cost (investments and operation).

What MT has to offer on the negative side:

- poor quality on big, expensive systems;
- very poor quality on cheap PC based systems;

but on the positive side we have:
- editing and checking tools;
- dictionaries and term banks;
- translation memories, i.e., systems capable of storing fragments of texts and their translations, with retrieval capabilities based on exact or fuzzy pattern matching.

The conclusion is that there is little reason for this group to resort to MT systems. The big ones are too expensive, and the cheap ones deliver a translation quality which requires so much editing that most translators would prefer to do the translations themselves rather than correcting the system's results.

Yet, in the area of tools and facilities, there are a number of products (already on the market) which help the translator to increase their productivity (e.g., translation memories, which help to avoid translating the same or similar fragments more than once) and the quality of their work (e.g., by helping to make consistent use of the customer's standard terminology).

Some typical examples
- Mono- and bilingual dictionaries on CD-ROM;
- Translator's Workbench 2 (Trados, Germany), including translation memory, terminology database, translation editor, terminology extraction, etc;
- Translator Work Station (CITI, Canada), including access to previously translated texts, word processing, terminology management, document comparison, etc.

2.3. Monolingual users

The third group we can identify are those users who want to be able to produce translations from or into languages they do not know themselves.

Their requirements are the following:
- low cost;
- easy to operate;
- translations should be usable as they are (i.e., no checking or post-editing).

What MT has to offer on the negative side:
- poor quality on cheap systems, not good enough to send out to third parties.

On the positive side:
- possibly good enough for internal use (e.g., relevance checking of incoming messages);
- modem or network access to professional translation services (which may or may not use MT).

Conclusion: MT is no real option for monolingual users if they need translations for communication with other parties (e.g., user documentation of their products, business letters), but network access to translation services may offer a solution. Most modern PC based translation systems offer good opportunities for relevance checking of incoming information, provided it is available in electronic form.

Typical examples are:
- PC based systems such as GlobaLink, MicroTac;
- Access to Systran via Minitel in France.

3. Current trends on the market

A number of interesting developments can be seen on the market, all of which will contribute to overcoming the language barriers in Europe:

(1) Big MT systems are becoming more cost-effective because of
   - better environments;
   - better integration of MT in, e.g., document production;
   - cheaper but more powerful hardware.

(2) Smaller systems will become available with high quality output for restricted tasks and domains, especially in combination with controlled languages supported by appropriate authoring tools.

(3) Small systems capable of generating multi-lingual text or speech output on the basis of tabular or other types of structured input (as opposed to free text input).

(4) Multilingual information systems based on spoken dialogues.

Typical examples:
- Translators work benches;
- Meteo (TAUM, Canada), weather report translation (the classical example);
- Patrans (CST, Denmark), patents translation; - Restricted English for translation and documentation purposes (Bull, Perkins Engines, Xerox);
- Multilingual traffic information via car radios (Bosch, Philips).

4. Trends in research

Strange enough, the trends in research as reflected by scientific journals such as Machine Translation and conferences such as TMI (Theoretical and Methodological Issues in MT) seem to have a fairly loose relationship with what is happening on the market.

As a matter of fact, where market developments tend to go towards more specialized and restricted systems, the general trend in research seems to be to widen the traditionally rather narrow horizon of the MT research community. Three main dimensions can be identified:

(1) From sentence to text: Discourse-based translation, essential for, e.g., pronominal reference problems.

(2) Looking beyond linguistics: Domain-based translation, where disambiguation takes place on the basis of domain restrictions.

(3) Looking beyond traditional methods: Combinations of traditional (rule-based) and statistical approaches, where preferred linguistic analyses and translations are chosen on the basis of statistical data, gathered from large mono- and bilingual corpora.

5. Conclusions for the TELRI project

Current trends in MT, both in research and on the market, show that there are a number of areas where a massive effort with respect to linguistic resources is needed:

- First of all, there will be an increasing demand for lexical and terminological resources in all Eastern European languages in order to serve as the basis for better tools for professional translators.
- Secondly, the definition of appropriate controlled or subdomain languages will only be possible on the basis of large collections of real-lifa data.
- Thirdly, the increasing popularity of statistical methods in MT will require the availability of more and more mono- and bilingual text corpora.

6. For more information

Hutchins and Somers. 1992. "An Introduction to Machine Translation". Academic Press, London; (encyclopaedic overview).

Arnold et al. 1994. Machine Translation, An Introductory Guide, Blackwell, Manchester; (introduction).

Machine Translation (editor Sergei Nirenburg). Kluwer, Dordrecht; (scientific journal).

Language Industry Monitor (editor Colin Brace). LIM, Amsterdam; (journal, with excellent market information).

European Association for Machine Translation. ISSCO. Geneva; (professional organisation of users and researchers).

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: TELRI - Proceedings of the First European Seminar:"Language Resources for Language Technology", Tihany, Hungary, Sept. 15 and 16, 1995	
Author(s): Heike Rettig (Ed.)	
Corporate Source:	Publication Date: 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

	The sample sticker shown below will be affixed to all **Level 1** documents	The sample sticker shown below will be affixed to all **Level 2** documents	
[X] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) ***and*** paper copy.	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but ***not*** in paper copy.
	Level 1	**Level 2**	

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at **Level 1.**

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature:	Printed Name/Position/Title: Norbert Volz, M.A. TELRI Project Manager
Organization/Address: Institut für deutsche Sprache R 5,6-13 - 68161 Mannheim Postfach 101621 - 68016 Mannheim	Telephone: +49 621 1581-437 / FAX: +49 621 1581-415
	E-Mail Address: volz(at)ids-mannheim.de / Date: 28/11/97

(over)