

Teaching Machine Translation in non Computer Science Subjects: Report of An Educational Experience within the University of Orleans

Viviane Clavier
Céline Poudat

Université d'Orléans
Faculté de Lettres, Langues et Sciences Humaines
Rue de Tour
45072 ORLEANS cedex2
e-mail : celine.poudat@univ-orleans.fr
Laboratoire d'Informatique Fondamentale d'Orléans
Rue de Chartres
45 067 ORLEANS cedex2
e-mail : viviane.clavier@iut.univ-orleans.fr

Abstract

Machine Translation is increasingly being taught within non scientific subject areas at French universities, which involves instructors solving educational and scientific problems caused by the lack of training of these students in computer science. Most of these students are being taught MT within the framework of language and linguistic courses.

As MT instructors in both Departments of Foreign Language and Linguistics at Orléans, we will report on our experience of teaching. Besides setting up the technological environment, we also had to consider the courses from two different angles. First of all, we can state that MT tools enable future users to enhance their skills in Machine-Assisted Translation, and secondly they introduce potential future system designers to computational linguistics issues.

Key-words : Teaching, Machine Translation, Machine-Assisted Translation, non computer science majors, computational linguistics.

Introduction

In French universities, Machine Translation (MT) is more often taught in Faculties of Arts – Linguistics and Foreign Language Departments – rather than in Faculties of Science¹ – Computer Science Departments. In Foreign Language Departments and especially in *Langues Etrangères Appliquées (LEA)*². In these courses, MT is taught within the context of translation theory, whereas in linguistic departments it is taught within the context of computational linguistics. In both cases, MT is tackled as one of the main domains of applied linguistics.

Indeed MT is a way of introducing students to the kind of skills that are used in the language industries and we are witnessing the emergence of an increasing number of courses on formalisms, computer science and applied linguistics.

In these non computer science subjects, the way of teaching MT depends on whether students belong to Foreign Language Departments or Linguistics Departments.

In LEA courses, MT classes are more often presented as Machine-Assisted Translation (MAT) courses covering “all techniques for automating the translation activity” (Boitet, 1994). Beside the other subjects connected to

office automation, courses aim at familiarizing future professionals with the working environment of a translator. MT courses are normally offered to postgraduate masters students (French Second Cycle³) following vocational courses leading to a professional qualification and to third cycle *DESS* students⁴.

In this confusing educational context, the teaching of MT is tackled from a pragmatic point of view, *i.e.* it is centred on the needs of the users and on the task of translation. Thus much of our work involves introducing students to practical work with tools (hardware and software). The

³ The French educational system is divided into three cycles :

- The first cycle (2 years of study) is open to the majority of graduates and leads to the *DEUG* (the French University General Education Diploma).
- After the *DEUG*, students enter the second cycle of their studies leading to the *licence* (equivalent to a B.A. in the U.K.) and the *maîtrise* (Masters)
- The third cycle offers two routes : a one year professional training program culminating in the diploma of higher specialized studies (*DESS*) and a research-oriented diploma which prepares to PhD.

⁴ Here are some CL-oriented *DESS* :

- *Techniques de la traduction et de la traduction spécialisée* (Toulouse 2),
- *Langues et Techniques* (Rennes),
- *Traductive et gestion de l'Information* (Paris 7),
- *Ingénierie multilingue* (Paris 7)
- *Métiers de la Traduction : Traduction éditoriale, traduction technique* (Aix-Marseille)

¹ To our knowledge, no MT courses are given in first and second cycles in scientific subjects even by the most famous French MT research center (GETA, Grenoble) – see footnote 3.

² *LEA*: combined honours course in Applied Languages and Business Studies.

type of tools used varies according to the translated languages, the covered domains, and to the quality of the translation. Users have to be introduced to the proper translation environment, manipulation of computer software, initiation into information retrieval systems, adaptation or creation of tools – the setting up of platforms of translation, creation or adaptation of multilingual dictionaries. Finally, introducing students to the workings of MAT involves familiarizing them with the wider dynamics of the translation process, and also learning them to know, for example, which computer program to apply to a particular translation task (pre-edition or post-edition stages).

In linguistics, MT is presented within the framework of *Computational Linguistics* (CL) as one of the major applications (T.A.L, 1992). Indeed, MT is closely connected to the historical background of computational linguistics (ALPAC report, 1966) and because of Internet growth, MT nowadays plays a key role in multilingual information. In this context, the performance of systems, architectures, dictionaries, grammars and formalisms has a center importance.

Even if the way of teaching MT is rather different whether you teach it in Foreign Language or in Linguistics Departments, instructors are faced with similar scientific and didactic issues :

- What is to be taught to students who have very little knowledge of computer science ?
- How can MT be integrated in non computer science subjects to fit the students' skills ?

As instructors at the Faculty of Arts within the University of Orléans, we were recently brought to introduce an initiation course in MT. Consequently, we had to face all these questions, teaching both in Foreign Language and Linguistics Departments. *i.e.* in *LEA Maîtrise* (second cycle) and in Linguistics *DEUG* (first cycle). Besides the different perspectives specific to each program, the level of the students also had to be taken into account in the introduction of MT teaching.

As the MT courses we teach are representative of those given in French Universities, it may be interesting to give an account of our experiences.

1 The Educational Context

In *LEA* courses, the training takes place within the framework of practical work (12 hours) and in *DEUG*, within lectures (4 hours) and practical work (4 hours).

1.1 Students' Profile

1.1.1 In *LEA*

The students in the *LEA* translation-oriented *Maîtrise* develop skills in foreign languages (in English and in another foreign language, chosen between German, Spanish and Japanese), in law and economics. A training in French grammar and stylistics is also offered. At the end of their studies, they will work as translators or they

will complete their training with a general or specialized translation *DESS* course. In each of the courses they follow, they will have to familiarize themselves with various forms of software, which explains the strong motivation and the high interest they have showed in the course. Finally it is worth pointing out that an introductory linguistic lecture on the theories of translation was given to the students during the first half of the academic year.

1.1.2 In Linguistics

The second year students have to attend a course on computer science which contains three modules : Office automation, Computer-assisted Learning (Presentation of educational reading software reading) and Computational Linguistics. It is within the framework of this last module which lasts 42 hours that MAT is introduced. Furthermore, the students receive theoretical instruction on general French linguistics (including phonology, morphology, lexicology, syntax and semantics, psycho- and sociolinguistics).

1.2 The Means at Our Disposal

1.2.1 The Material Environment

The Faculty of Arts in Orléans owns two classrooms each fitted with 15 PCs which are all Internet-connected *via RENATER (REseau NATional d'Enseignement et de Recherche)* Beside these two classrooms, a multimedia room is available at any time. Moreover, the books mentioned in the bibliography we gave them at the beginning of the year are easily accessible in the University Library.

1.2.2 The Teaching Equipment

The time allotted to courses in CL and MAT is weak in comparison with the time given to other courses in the *Maîtrise* and in *DEUG* syllabus. Consequently, it was difficult to solicit funds to invest in the purchase of software programs specialized in translation. In these conditions, we had to resort to **online tools available on the web**. However, financial assistance may become available in the academic year 2001-2002, as a new *Specialized Translation, Culture and Media DESS* has just been created.

1.3 Objectives of the Course

1.3.1 In *LEA*

The main objective of the MAT course is to familiarize the students with the working environment of the translator : on the one hand, the evaluation of software, the role of the translator, *i.e.* deciding when human intervention is necessary in the machine translation process. On the other hand, the evaluation knowledge and evaluation of the online MT tools on the Internet (dictionaries and MT software programs). We concentrate

on both translation from English into French and *vice versa*.

1.3.2 In Linguistics

CL classes in linguistics aim to introduce the students to computer science and formal linguistics, using an inductive approach. In practical work classes, they learn how to use MT tools and the Frantext⁵ textual data basis. Then, they have to analyze a specific linguistic problem from a formal point of view.

2 The Skills Taught In Mt Courses

2.1 Technical Skills

2.1.1 Office Automation Applications

Generally speaking students are not competent in the use of Microsoft Office tools (Word, Excel and Access data bases) but nevertheless, the *maîtrise* students will have to use these tools in their future profession. That's why they are encouraged to work jointly with the Internet and on Microsoft Office tools. For instance, they have to draw Excel spreadsheets listing various mistranslations made by MT software programs.

2.1.2 The Internet Network

An introduction class to the Internet network and to search engines turned out to be necessary when we set up this course. The advanced functions of the search engines (Boolean queries and keywords) allowed them to be more efficient.

2.1.3 Mt Software Programs

Thanks to this computer science initiation, the students are able to list and to use online machine translators (powered by Systran, E-lingo, Prompt/Reverso, Intertran and TranscendRT), dictionaries (Logos, Wordreference, Research-it!, Eurodicautom, *etc.*), parallel texts alignment (Rali), tagging and parsing software (Latl, Winbrill and Xerox). No downloadable tool is used.

2.2 Theoretical Skills

2.2.1 Linguistic Skills

In linguistics, the lecture contains four parts : the first part deals with the historical background of multilinguality-bound MAT and its economic and political issues.

The second part aims at describing the sub-areas of the language that can be automatized in the translation process. For example, the regular French-to-English and English-to-French examples (place of the adjective in the noun phrases) illustrate the legitimacy of a possible computer treatment. Then, examples which prevent

formalization are given: polysemous words, references of pronouns, problems of tense and aspect in the translation of the French present.

The third part of the course introduces both fundamental techniques of MT (transfer and interlingual approaches), with examples of adapted formalisms to lexical and syntactic transfer.

2.2.2 Linguistic Evaluation Of Mt Tools

All students working with MT/MAT must be able to evaluate the tools with which they are working.

As we do not have access to the internal architecture of the MT systems, we can only evaluate the tools from a linguistic point of view. Thus, we have so far focused on the linguistic evaluation possibilities of the system, which are all based on machine translations.

Several evaluations are currently used ; one of them is based on a typology of the mistranslations we imagine a translator is able to make. Anne-Marie Loffler-Laurian used such a typology to evaluate Systran (Loffler-Laurian, 1994) and the students were asked to build such a grid. The weaknesses of the linguistic treatment were then pointed out. For instance, many errors in determination reflect a weakness in the linguistic treatment. As we roughly knew the general architecture of the online translators (the Systran system is for instance described on the site and many Systran evaluations have been made), it was possible to 'reconstitute' the linguistic treatment.

Nevertheless, such an evaluation is not representative of the intelligibility of the translation, as many errors do not affect it (grammatical words in general). Another evaluation is then needed to calculate the intelligibility rate of the text : a semantico-pragmatic analysis. Finally, it is also possible to evaluate the fidelity of the target text to the source one.

Such evaluation methods have been approached in the MT courses we have so far given: knowing how to evaluate MT tools is essential.

2.3 Evaluation of the Students

2.3.1 The Examination Question

In *Maîtrise*, the examination question this year was divided into two parts : a French-English translation and an analysis question. In the first part, three different genres of texts were translated : a technical document, a scientific document and a legal text. In the second part, the students were asked to analyse two sentences as a transfer-based MT system would do. Then, they had to both underline the flaws of such a linguistic treatment and to propose improvements.

In *DEUG*, the examination question was presented within the framework of a newspaper article attacking MT. The article questioned the quality of MT systems. The students were asked to look for the online tools that may have generated the translation quoted in the newspaper article. Then, they had to give an evaluation of these tools

⁵ Frantext : information retrieval system based on textual data basis. <http://www.inalf.fr>.

depending upon how many errors (and what type of errors) they found.

2.3.2 Analysis of the Results

In *LEA*, the students described the translation problems they had come across in each of the three texts (problems of understanding due to the specialized vocabulary; syntactical, morphological or lexical problems, *etc.*). Then, they focused on the task of translation, using the online tools they had listed during the class. As they didn't have a real environment of MAT, they had to reconstitute it with the means they had at their disposal. Therefore, the way they proceeded was also tested, and the results were quite satisfactory. Finally, they had to propose different methods to improve the linguistic treatment of the transfer model, and it is worth noting that their suggestions were very diverse: propositions of statistical models, new labelings of the constituents of the sentence, use of tree-like structures, *etc.* They were however not very realistic.

3 Appraisal and Future Prospects

3.1 Common Remarks Concerning Both Subjects

3.1.1 Lack of Relation Between Theory and Practice

Broadly speaking, the students do not perceive the relation between the theoretical presentation (transfer/interlingual approaches, organisation of the dictionaries, role of grammars...) and the online tools. Intuitively, the *DEUG* students punctuated their criticisms with remarks concerning the stupidity of the computer, the fact that it was "silly"; and however, in the *Maîtrise*, the exercise in which the students were asked to make the various stages of analysis of the MT system appear (on the syntactic and lexical level) was not very convincing.

Bearing this in mind, an educational software program which shows the various stages of the process of translation in a clear fashion would be welcome. However, at this point in time, it is hard to imagine this.

3.1.2 Comments on the Material Environment

The inventory of online freeware tools takes up a large part of the students' time. Of course it leads to the listing of a large number of MT software programs, but the same systems are very often accommodated by sites with different names and it is quite confusing. What is disturbing is that two translating systems with different names produce the same translation to within about an error: for example, Worldlingo produces curiously similar translations to those of Systran.

3.2 Remarks Peculiar to the Specificities of the Two Subjects

It is also interesting to point out that our respective students don't show the same motivation. Students in *LEA* enter professional world fairly quickly. That's why

they show a lot of interest in searching for tools which will assist them in their future tasks. The motivation of the *DEUG* students is lower, thus it is necessary to prepare them more for the role of the linguist who participates in the design of software.

It is necessary to bring about changes in the teaching of MT because more and more linguistics courses are CL-oriented. MT represents an important opportunity for these students. Developing more partnerships among linguists and computer specialists is fundamental; as a matter of fact, knowing exactly what computer specialists expect from linguists and conversely would be required. The CL specializations, and the specializations in computer science generally begin in the 2nd cycle of studies, except for a few cases (for instance, a new *DEUG* at the *Sorbonne* University aims at developing two skills: languages and computer science). Thus, it is necessary to train the students quickly in order to make them operational and to make their entrance into the job market easier.

Conclusion

An MT (or MAT) course definitely plays an important role in non computer science subjects, in both Foreign Language and Linguistics Science Departments. MT enables the students to unify the skills they have developed throughout their studies. Indeed, the *LEA* students are led to use both their language and linguistic skills, whereas the Linguistics students have the opportunity to link all the theoretical subjects they learned earlier on (like semantics, syntax, morphology, *etc.*). MT courses help the language students to master their own translator environment and the linguists to understand their role in computational linguistics.

The teaching of MT has to be clearly thought out, as instructors have to familiarize the students with the machine, which is often perceived as frightening and autonomous. Explaining how the computer works is often necessary. Furthermore, an MT course is more formal than a language course; technical details concerning the system architecture are given, and formal linguistic models are presented. Consequently, MT instructors should never forget to link theory to practice. Otherwise, the students may be seriously confused. In this respect, practical work is fundamental. Besides, this is something that has been underlined in the survey into tools and techniques for teaching Machine Translation (1996) concerning the teaching of MT. Most instructors expressed their wish to have access to MT Software and many underlined the importance online tools may have in MT teaching.

In our own experience, we have found these tools to be very useful: on the one hand, because they introduce our students to the Internet, and on the other hand, because they can be utilised in a variety of different linguistic contexts. The practical work in our course is designed to get the students used to working with MT and so far we

have been successful in this. However, the introduction of interlingual MT systems or translation-memory-based software would be welcome.

- TRASBORG, Anne (Ed.) (1997). *Text Typology and Translation*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

References

- ABOU, Thierry and LEFAUCHEUR, Isabelle (1987). *Les Industries de la langue*. Daicadif, volumes I et II.
- ALPAC (1966). *Language and machines ; computers in translation and linguistics*, Washington D.C.
- ARNOLD, D.J. *et al.* *The MT book (1994)* : <http://clwww.essex.ac.uk/MTbook/PostScript/>
- BALKAN, Lorna, DOUG, Arnold and SADLER, Louisa (1996). *Tools and Techniques for Machine Translation Teaching: A Survey*. <http://clwww.essex.ac.uk/group/projects/MTforTeaching/>
- BOITET, C. (1994). (Human-Aided) Machine Translation : a better future ? : <http://cslu.cse.ogi.edu/HLTsurvey/ch8node5.html>
- GRISHMAN, Ralph (1986). *Computational linguistics, an Introduction*. Cambridge University Press.
- ISTVAN S. Batori (Ed.), LENDERS, Winfried (Ed.) and PUTSCHKE, Wolfgang (Ed.) (1983). *Computational linguistics/Computerlinguistik, An International Handbook on Computer Oriented Language Research and Applications/Ein Internationales Handbuch zur Computergestützten Sprachforschung und ihrer Anwendungen*. Walter de Gruyter, Berlin, NY.
- LEHRBERGER, John and BOURBEAU, Laurent (1988). *Machine translation, Linguistic characteristics of MT systems and general methodology of evaluation*. *Linguisticae investigationes*: supplementa, *Studies in French&General Linguistics/Etudes en Linguistique Française et Générale*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- LOFFLER-LAURIAN, Anne-Marie (1994). *La Traduction automatique.*, Presses Universitaires du Septentrion, Villeneuve d'Ascq.
- NEWTON, John (1992). *Computers in translation, a practical appraisal*. Routledge, London and NY.
- RIAO 97 (.25-27th June 1997). *Conference proceedings with Prototype and Operational Systems Demonstrations*. Computer-assisted information searching on Internet, MacGill University, Montreal, Quebec, Canada.
- T.A.L. (1992). *Traitement Automatique des Langues*. Spécial Trentenaire, vol. 33, Numéro 1-2, Paris.
- T.A.L. (1996). *Traitement Automatique des Langues*. L'enseignement du TAL, vol. 37, Numéro 1, Paris.