

Teaching Machine Translation evaluation by assessed project work

Judith Belam
Foreign Language Centre, University of Exeter
Queen's Building
The Queen's Drive
Exeter EX4 4QH
UK
J.M.Belam@ex.ac.uk

Abstract The paper describes the use of an assessed independent study project on machine translation evaluation as part of a final-year undergraduate course on machine assisted translation. The advantages and potential drawbacks of the use of such a component are outlined. The usefulness of studying MT evaluation is underlined, as it includes consideration of many aspects of MT. The suitability of teaching MT to language learners is considered, as are the transferable skills students can hope to gain after following the course.

Introduction

The course, entitled "Machine-Assisted Translation" (MAT) is for final-year undergraduates in Modern Languages. It runs over one semester and is worth 15 credits of a 120-credit study programme. Students will have achieved a good working competence in at least one foreign language and will have spent a year in the target language country. They may have used computer-assisted language learning materials but no knowledge of MAT tools is assumed before they start the course.

The MAT course is not primarily intended as a language teaching course. Students will continue their language learning concurrently on a course consisting of oral classes, essay writing and short literary texts for translation into and out of the target language. The MAT course on the other hand aims to provide a broadly based introduction to MAT including the basics of text processing, practical aspects like pre- and post-editing

and dictionary creation, and the history of MAT.

The assessment of the course falls into two parts. 50% of the marks are given for performance in a 2-hour written examination and 50% for an independent study project. The independent study project consists of an evaluation of one or more MT systems. The guidelines given to the students are as follows:

The portfolio takes the form of a report on an evaluation of a machine translation system, which you will design and carry out. You will be studying the subject of machine translation evaluation formally in weeks 6 and 7, and you will then be ready to proceed with the preparation of your portfolio. It is expected that you will begin with a short introduction outlining the principles of machine translation evaluation; this will include types of evaluation and questions of why, by whom and for whom evaluations may be carried out. You will then describe the type of evaluation you yourself have decided to do: you may be focussing on output (accuracy, readability, coherence ...) or on the system itself (ease of use of dictionaries, number of additional features available, handling of text within word processing programs ...); or you may have decided to look at the way the program handles specific problems (subject domains, or specific grammatical structures); or you may address a practical problem like how much texts need to be pre- or post-edited in order to arrive at an acceptable translation. You will give your criteria,

explaining why you have chosen them, and describe the way in which you have decided to test the system. Your evaluation may be comparative (Systran vs. Globalink) or may focus on one of the systems. Finally you will report on the process of your evaluation, give details of the tests you used and how you arrived at your conclusions.

The whole portfolio will not exceed 3,000 words; this does not include any examples you give. Machine-translated texts and their originals are not included in the word count and should be attached in a separate appendix.

As you will see from the programme there are several sessions devoted to supervised work on this assignment and it is expected that you will discuss your methods and progress with the tutor in the early stages.

Students therefore design and carry out their own mini-evaluation. Examples of some subjects which were chosen were:

- terminology and dictionary tools in Systran and Globalink
- the translation of phrasal verb constructions
- comparative evaluation of Systran and Globalink's translations of children's non-fiction
- comparative evaluation of the translation of jokes

Supported self-study

There is an inherent contradiction in the concept of a self-study project which is to be assessed and will contribute to the final mark gained by the student. For it to form part of the students' learning the tutor must provide guidance and feedback. On the other hand, for it to form a fair part of the assessment, it must represent the student's own unaided work and independent competence. We have tried to satisfy these conflicting requirements by providing extensive support while the projects are in preparation, but leaving the student to produce the final version on their own. In practice students are well

aware of the unwritten rules governing this type of project and rarely seek to gain advantage by asking for too much help in the later stages.

Why teach MT evaluation?

Full-scale MT evaluation is a specialist field clearly beyond the scope of a short university course. Students are not expected to do extended research into the field itself, but have two lectures outlining the basic principles and are given some additional reading (Hutchins and Somers 1992, Trujillo 1999, Somers in press). Within the limited scope of their projects, and working entirely in the abstract, it is unlikely that they will arrive at practically useful conclusions about any particular system. However, the value of including a study of evaluation in the course is 3-fold.

(i) it is more than likely that language graduates will be expected to consider the use of MT systems at some stage in their career. Having studied MT evaluation in this way, they will be equipped to provide a sensible and realistic answer to obvious questions like: "Is MT any good?" "Can it save us money?" "Which system should we buy?" – questions which are easily asked but not so easily answered.

(ii) Evaluation of MT output requires students to take into account many other areas covered during the course. They are encouraged to consider not just the raw output but related questions like the amount of pre- or post-editing that is necessary or the impact of using the dictionary tools. In this way the project constitutes an important stage of their study of MT, when they must bring together all they have learned and consider how it is put to use. The project is marked according to the following criteria:

1. Knowledge of the principles of evaluation
2. Awareness of how project chosen fits into the context of these principles
3. Realism of project

4. Appropriate development of evaluation method, e.g. creation of scale
 5. Choice of source materials appropriate for MT
 6. Awareness of characteristics of texts/materials chosen
 7. Rigorous performance and analysis of evaluation
 8. Critical assessment of project and suggestions for improvement
- (iii) Finally, it obliges students to examine their preconceptions and their first impressions of MT. These may vary widely (Gaspari 2001). Some students, comparing instinctively with the careful manual translation which their concurrent language courses require, tend to be very dismissive of raw MT output. Others, who may already use it for translating material online, tend on the other hand to overestimate its capabilities. At least one student was rash enough to confide that she was in the habit of using Babelfish for a first draft for her literary translation work! Work on MT evaluation must be set in some sort of context, and this forces all students to realise that the question "How good is this translation?" is just about as useful as asking "How long is a piece of string?" They must define in what circumstances the translation is likely to be made, for whom and why.

The success of the independent project component

The exercise has proved to be a very valuable one. In my experience an assessed self-study component in general is an extremely effective way of motivating students. Some reasons for this are rather negative: they cannot leave the work until the last minute, or leave out vital areas in the hope that the exam will not test them on a particular aspect. However there are also more positive advantages.

Requiring students to choose their own topics has obvious benefits. They tend to become more interested in a subject which they have chosen

themselves, and in which they have invested a lot of time and effort. They are more engaged with the project from the outset; they remain more committed; they work more independently. They are less likely to become bored, or crib from textbooks or from each other. A further benefit is that they will inevitably come up with ideas the tutors would not have thought of, thus broadening the whole scope of the course.

In some cases the process of choosing the topic turned out to be a useful learning experience on its own. Some students started on one area, only to discover unexpected problems or difficulties which necessitated a change of direction, and often this change taught them as much as the subsequent work on the project. Some, for example, arrived at a better understanding of the difficulty of arriving at an absolute standard for evaluation when they discovered that it was easier to compare two systems than to concentrate on just one. Another candidate began looking at children's non-fiction, to realise that the translation of entire books by MT would be an unlikely use for the system. He refined the scope of his investigation so that his source texts consisted of short passages suitable for introducing displays designed for children in exhibitions or museums. Within this more practical context he was able to go on to produce a more reliable evaluation, and he arrived at a more realistic appreciation of the contexts in which MT is likely to be used.

Some difficulties, however, also arose with the choice of topic. Some weaker students had enormous difficulty choosing a topic and required very detailed guidance, which raises questions about the fairness of the exercise, given that it counts towards the final mark. It is hard to take into account objectively the amount of help given to an individual. Secondly, some chose topics which were much more difficult than others. The candidate who chose to look at the

translation of humour plainly had an enormous amount of background research to do on the subject of what constituted a joke, how far the humour lay in the language and how far in the subject matter, cultural background, personality of the reader, etc. Compared with this, the student who decided to compare the usefulness of the dictionary tools in Systran and Globalink had a relatively easy ride. Once again this proved difficult to take fairly into account when assigning a final mark. Finally, and this is a difficulty related to teaching MT evaluation in general, some students had trouble limiting the scope of their enquiry. Sometimes this was simply due to an inadequate appreciation of the whole context in which the evaluation should be considered. One very general project, for example, simply took a series of different texts, including extracts from a children's fairy tale, a videogame instruction manual, a commercial order form, and a medieval epic poem. Needless to say 3000 words were insufficient to cover all the issues raised. Sometimes the problem arose because students had not understood the level of detail required in their analysis. Another candidate attempted to analyse passages with specialised terminology taken from twelve different subject domains and translated into both French and German. My advice to reduce this enormous task resulted only in a reduction to seven topics and the resulting project contained more description than evaluation and analysis, as the student was overwhelmed with the sheer volume of material to be examined.

Is the study of MT evaluation suitable for language learners?

This course does not set out to be a language course as such. It assumes a good level of knowledge of the foreign language in order to learn about MT. However, the students are all language learners and it is therefore important to

consider the suitability of the subject. Two main areas need to be addressed:

(i) It is important to consider whether it is appropriate to expect language learners to give an opinion on the relative quality of a given translation. In all their other language work the greatest care is taken not to expose them to incorrect models, and it is the role of the teacher to assess quality and to reject anything which is inadequate. Work on MT evaluation, on the other hand, not only assumes that the learner is competent to give an opinion, but also expects him to give a relative judgment, and to accept as adequate for a given circumstance a version which would otherwise be thrown out as "wrong".

I believe that while this difficulty must be recognised, it should not be overstressed. Firstly, the whole concept of "fit-for-purpose" translation, discussed in more detail below, is so important that it must override the principle of not using incorrect models in teaching. Secondly, students are more than capable of making these distinctions for themselves. After half an hour spent trying out the systems they will be well aware of the difference between raw output and high-quality human translation, and so long as this difference is constantly borne in mind the "damage" done by exposure to errors will be limited.

(ii) It is perhaps more important to ask whether teaching MT evaluation is actually helping students' language learning. In theory, the course is not specifically designed to improve students' language skills. Furthermore, students' competence in the language does not form part of the assessment criteria, especially as the course is designed to be non-language-specific.

In practice their competence does of course affect the quality of the work submitted as part of the portfolio. Students frequently expressed the concern (in some cases with a certain amount of justification) that they were not able to

judge the appropriateness of a particular version because they did not feel they had a good enough appreciation of the language. In this way the course can in practice contribute towards improvement of their knowledge of the foreign language. A consciousness of any gaps or uncertainties can only be useful, and did appear to lead in many cases to the kind of intensive dictionary and grammar work which classical translation courses aim to stimulate.

Transferable skills

When the course was first set up considerable attention was paid to the skills students could expect to acquire which would be of use to them in areas of activity outside the field of MT (Belam, 2001). The independent study project has indeed furthered the acquisition of some of these skills, although students are sometimes resistant to being expected to improve their competence in areas they do not see as specifically related to the course (*"I didn't come here to improve my English"* was a complaint heard more than once, even though a thorough competence in the mother tongue is an essential prerequisite to being a good translator). In particular students gain a broader understanding of an area which is usually formally avoided on conventional language courses, that of the value of imperfect communication. I have described how, when preparing their projects, they are encouraged to consider the type of texts they are using, and in particular the circumstances in which they will be translated. The quality of the translation is then assessed not in the abstract but in the particular conditions which have been specified, and a relatively poor translation can sometimes be seen as perfectly adequate for a particular purpose. One candidate, for example, took online newspaper articles as his source material and showed a good understanding of the quality of translation which would be required, explaining: *"In*

the case of a newspaper report which just reports facts and not feelings, the loss of the writer's style or register is not that important. The sense and meaning of the article should still be portrayed when the text has been translated". This awareness of the "fit-for-purpose" translation is a valuable element of the course.

There have also been some unexpected additional gains which have come out of the course. The projects were demanding and complex to design and set up, and several students underestimated the time it would take to assemble, analyse and present their findings. This appreciation of the importance of the "writing-up stage" will be important to students taking up any form of research or report writing. Several of them were introduced to the use of metrics and statistical methods, and some of the issues associated with interpretation and presentation of the figures they produced. They were all brought to consider the importance of the computer/user interface, and the value of understanding some of the workings of the system in order to get the best out of it.

The most important unexpected effect of doing the course, however, was what one could call an increased linguistic awareness. At least one student said to me partway through the course that he found that the course was having a distracting effect on his other study, as he had got into the habit of noticing linguistic features of what he was reading which would make a text suitable or unsuitable for machine translation. *"It's a whole new way of looking at language"* were his words. He was almost complaining at the inconvenience, but at the same time he realised that he was in the process of gaining an whole new perspective on language.

Conclusion

The inclusion of an independent self-study project on evaluation in the MAT course has proved to be a very valuable

aid to students' learning. It is a demanding exercise which furthers understanding not only of evaluation techniques but of MAT as a whole.

References

- Belam, Judith (2001). Transferable Skills in an MT course. *MT Summit VIII Workshop on Teaching Machine Translation*, Santiago de Compostela, pages 31-34.
- Gaspari, Federico (2001). Teaching Machine Translation to Trainee Translators: a Survey of the Knowledge and Opinions. *MT Summit VIII Workshop on Teaching Machine Translation*, Santiago de Compostela, pages 35-44.
- Hutchins, W.J. and Somers, H.L. (1992) *An Introduction to Machine Translation*. London, Academic Press.
- Somers, H.L. (in press) *Computers and Translation: A Handbook*, to be published by John Benjamins; copy made available in draft form by kind permission of the author for consultation by our students. On evaluation, specifically the chapter by John White.
- Trujillo, Arturo (1999) *Translation Engines: Techniques for Machine Translation*. London, Springer.