Meeting of the Chairpersons of the Committees on Education, Science and Culture and the Committees on the Development of the Information Society

Session I: Machine Translation of European Languages: Problems, Challenges, Prospects

Machine Translation: The Virtuous Circle - Presentation by Maria Cristina de Preter, Portuguese Language Department, Directorate-General for Translation, European Commission

Good morning, ladies and gentlemen, *(MT: Labas rytas, ponios ir ponai.)*

It is an honour to address you at this place and time.

This is just an example of how useful Machine Translation (MT) can be. But don't get overoptimistic. As you probably know, MT is not always that good.

My name is Cristina de Preter and I am Head of the Portuguese Language Department, which is one of the 24 Language Departments in the Directorate-General for Translation (DGT) of the European Commission.

I have been for many years involved in the operational side of translation and I started using Machine Translation in my own translation work more than 10 years ago.

Therefore I have a hands-on and an almost "historical" view of MT use within DGT, both with the previous rule-based system and now with the new statistics-based open-source system *Moses*, which is being developed within EU-funded research projects.

I was asked to present MT@EC – the new MT service – and I would like to do so mainly from a translator's perspective. But first, let's just have a quick look at the larger picture.

Language diversity has been a building block of the European Union. Fortunately, the situation in Europe is not as severe as at world level, where – from the 6000 reported languages – about 40% are vulnerable or endangered.

With 24 official languages – and this number is likely to increase with further enlargements – and about 60 regional languages, the EU faces a challenge to have an inclusive but sustainable language policy. And we are committed to meeting that challenge.

MT in the globalisation era

So the question is: **Can machine translation help? And if so, how?**

Let me give you my perspective as to the contribution DGT – and its Machine Translation service MT@EC – can give to multilingualism and a sustainable translation policy in the global society.

In the globalisation era, translation – including Machine Translation – is increasingly becoming a utility.

Machine Translation has been made freely available on Internet search engines and is now widely used. We all know MT services like *Google* and *Bing*.

The fact is that there is a fast-growing need for the translation of all sorts of documents – from EU legislation to e-mail messages (in social networks) – and with different quality/reliability requirements.

This need cannot be satisfied by "traditional human" translation alone, as we would need to have half of the population translating for the other half… or we would have to opt for a single language, thereby contributing to the "disappearance" of languages that have a relatively small number of (native) speakers. This would go against everything Europe stands for.

Therefore, Machine Translation has an important role to play.

European Commission – a major actor in this field

The European Commission has been a major actor in the field of Machine Translation, both as a long-standing investor in this technology for more than 35 years and as a regular user of MT in its translation workflow since the 90s.

The Digital Agenda's vision for 2040/2050 is that anyone will be able to access digital resources and interact with others in their own language, and that companies will be able to operate across borders without hitting the language barrier.

The Digital Agenda is being implemented with initiatives like the Framework Programmes on Research and Technological Development, the Connecting Europe Facility (CEF) and the initiative Interoperability Solutions for European Public Administrations (ISA).

EU-funded projects like *Euromatrix(Plus)/MosesCore* are of particular importance to Machine Translation, as they are a success story: many of the MT systems now available are based on the Moses "approach".

Furthermore, the *Horizon 2020* – the new Framework Programme on Research and Innovation – is expected to increase its funding in the area of Content Technologies and Information Management, which includes Machine Translation.

You will hear more about this subject in the afternoon session from my colleague Kimmo Rossi of DG CNECT.

MT@EC

So, what is the role of DGT in the EU far-reaching and multi-billion Euro Digital Agenda strategy?

The European Commission decided to promote MT in order to support some of its major policies, for example completing the Internal Market. DGT was charged with the mission of creating and running the MT@EC service to satisfy the needs of the EU institutions, of initiatives like Internal Market Information system, e-Health and e-Justice and of the Member States' administrations, which are all tools to better include citizens and provide them with better public services.

But, more specifically, when speaking about Machine Translation, we have to think about the practical needs it is meant to satisfy.

Is it just to have an idea of what a text is about? Then MT by itself or with some light post-editing may be enough.

Or is it for dissemination purposes, which means translation of publishable quality? That is mainly the work in DGT, and for that purpose, Machine Translation is just another computer tool available to translators.

For different needs, different solutions!

The MT@EC service that DGT provides aims to satisfy both needs, providing translations among all the EU official languages, which means 552 language combinations … although with different degrees of "performance".

A question often asked is: Why not using MT systems/services already available, such as *Google* or *Bing*?

The answer is:

➔ **Security and copyright:** Because EU institutions have confidentiality and Intellectual Property requirements. This concerns both the texts which are being translated as well as the corpora used to build the MT systems - called MT engines - since not all documents are (or can be) made publicly available, nor can copyright be a matter of dispute.

➔ **People:** Because DGT has the necessary linguists, translators and computer and computational linguist experts to tailor the service to its users' needs. This is a major asset.

➔ **Data:** Because MT quality cannot be judged in absolute terms. Its performance/usefulness depends not only on the language pair, but also on the documents to be translated and on the data the system is fed with. And therefore, the wealth of data the Commission has is another major asset.

Furthermore, DGT is also contributing to the promotion of MT by making freely available its public corpora in an aligned format suitable to train MT engines.

MT@EC and national administrations

Public administrations are already able to use MT@EC through online services like the Internal Market Information system.

DGT is now working on setting up pilot projects in which Member States' administrations will be invited to participate on a voluntary basis and without any cost on their side, other than the use of their internal resources, and without any commitment from DGT on the use of the service after the end of the pilot project.

The ultimate objective of MT@EC is to facilitate cross-border exchange of information between European public administrations by overcoming language barriers. MT@EC is to be a building block for new user-centric European public services.

Cooperation with national administrations can be fruitful for both sides, since with more national (and high quality) data, MT@EC quality could improve and better serve the needs of EU institutions and Member States alike. By data, I mean bilingual and monolingual corpora.

Moving further to events where it is a co-organiser or participant, DGT is also organising a series of annual conferences to promote MT. The first, held in 2012, targeted the translation services of the EU institutions (MT@Work) and the second, to be held on 6 December 2013 in Brussels, will target MT users in general (non-translators) in the EU institutions.

The third, to be held in November 2014, concerns the Member States' administrations. The conference is titled "European Commission Machine Translation: Supporting Multilingualism in EU Member States".

Please include it in your busy agenda. 🙂

Translator's perceptions about MT and the Virtuous Circle

Now, speaking from a translator's perspective, the reactions to Machine Translations are very contradictory. According to a survey made last year in DGT, some translators love it, some hate it and some others have mixed feelings.

For translators, Machine Translation is just another tool – an important one – that can help them do their job in a more productive way.

In the last 15 years, translators have been using computer tools that allow an easy reuse of previously translated sentences (the so-called Translation Memories) or the search of terminology databases like IATE. Now we are adding Machine Translation to the toolbox.

But translators can also help improve the quality of Machine Translation output. Both indirectly, by producing high-quality translations that are used to train MT engines and directly, by giving detailed feedback that may be used to improve the engines. After all, translators – more than any other users – are the translation "experts".

I like to think of this as the Virtuous Circle, with MT helping translators and translators helping MT. It's clearly a win-win situation and there is a place for all!

MT Problems

But – and of course there is a but – MT performance is not equal for all the language pairs.

For Romance languages, like Portuguese, MT is very useful, while for highly-inflected/morphologically-rich languages like Lithuanian, its degree of usefulness is typically (much) lower.

I imagine you have already tried MT into Lithuanian and seen the results. Sometimes it is just plain rubbish … but sometimes it helps (a lot)!

MT Challenges and Prospects

Machine Translation is now a very active area of research. Statistical MT has been a major development but some of its shortcomings are becoming quite apparent. For more complex languages and for better control of its output, rules and linguistic information will have to be added.

That's the next big thing: the much-awaited hybrid MT, with the promise of substantial improvements in MT quality, something which is particularly important for morphologically-rich languages or languages with insufficient data to feed the SMT engines. That is the case for new EU official languages and also for regional languages.

It is also a challenge to provide adequate information on MT achievements as well as its limitations to users in general as well as translators and translation students and raise awareness among them, so that prejudice and fears … or wild expectations … do not prevent the use and quick and reasoned uptake of MT.

So, in a nutshell:

- It is worthwhile investing in Machine Translation … even for language pairs for which MT results are not that good right now.

- It is worthwhile promoting cooperation at all levels.

  Besides the cooperation among EU institutions, it is very important to create synergies for cooperation between the Commission/DGT and:

  - national administrations;

  - academia, both in the areas of Language Technology – which develops applications – and Translation Studies – that train new translators.

    And here it is worthwhile noting the European Master in Translation and the Visiting Translator Scheme that reinforce cooperation between DGT and the outside world.

  - language industry.

Information on MT@EC

For detailed information on the MT@EC service, see the issue on Machine Translation, published by DGT in February 2013, which presents an overview of what is being done from the perspectives of IT experts, MT researchers, managers and translators.

In DGT Magazine *Languages and Translation*

(http://ec.europa.eu/dgs/translation/publications/magazines/languagestranslation/documents/issue_06_en.pdf).

Thank you very much for your attention.