

Institute of Translation and Interpreting
ITI 2003 Conference
12 September 2003 – University of Warwick, UK.

Translation Memory Debate

Panel Members

Joanna Waller – Freelance Translator & Member of the ITI Council (Debate Chair)
Michael Benis – Freelance Translator, Journalist & Copywriter
Bob Clark – Senior Lecturer, University of Leeds
Charlotte Couchman – Freelance Translator
Gerald Dennet – Managing Director, Star UK Ltd
Daniela Ford – Freelance Translator
Rob Giles – Senior Support Engineer, SDL Desktop Products
Bjorn Harris – Support Services Manager, ATRIL Software
Professor Tony Hartley – Director of the Centre for Translation Studies, University of Leeds
Marc Prior – Freelance Translator

Brief biographies can be found at the end of the report.

Debate

Joanna Waller

Perhaps we could start with the principles of using Translation Memory (TM) and what people think about it.

Philip Boyden

What is the world market for TM, what drives it and why are bugs and lack of features not corrected in successive versions of the products?

Gerald Dennet

Nobody knows how big the market is. The last figure I had for our own product, Transit, was 10,000 licences worldwide. The development of Transit is driven by the needs of our major translation customers in the automotive market. The only development, which is a response to the software customers, is the push to include right-to-left languages.

Michael Friedli

So, Star does not regard the freelance translator as a customer?

Gerald Dennet

Clearly we do. We also use Transit as a marketing tool to sell our translation services.

Michael Benis

It costs a lot of money to develop TM software and even some of the largest companies are working with reduced development budgets. If they do not think about the end customers, the translation agencies and the freelance translators, they are doing a disservice to all three. There has been a big focus in the last few years on developing systems that can interface with large web-enabled central databases, so that all the translators on a team are working from a common base. However, everyone will suffer if these systems are not robust and reliable. Developers know this and are concerned to address problems. This is a complex and expensive process and there will always be improvements that could be made. Users should always report problems.

Daniela Ford

It is all about building good relationships with software suppliers. They need to know what translators and clients need.

Henk Hazelaar

What can TM do for me? I do not do the repetitive jobs, where TM is really useful. However, I would like something that would help me to archive my information and get it into a more workable form.

Marc Prior

It is easier to generate an archive that you can access if you do it as you go along. It is probably not realistic to import all your legacy data into a system.

Michael Benis

It depends what you want to do with it and what printed format it is. One of the attractions of TM is that it will build an archive, but only for electronic source material. The problem arises when it is on paper. There are three approaches to this: retyping into an electronic format, speech recognition and Optical Character Recognition (OCR). However, OCR is only viable when the source documents are really high quality. Realistically there is not much you can do with the legacy material, apart from organising it so that you can find what you need. What you should do is to decide how you want to move forward, either with a TM product or simply using 'search files' on stored texts.

Marc Prior

A number of people impose a heavy surcharge for hard copy and that has resulted in far more work arriving in electronic format.

Henk Hazelaar

I usually translate patents, which arrive in paper form.

Charlotte Couchman

I also translate patents and give a small discount for electronic documents, which has produced some improvement. We scan any paper document that is more than six pages in length and have produced excellent results. The time to scan is offset by savings in typing time. There is a certain amount of repetition within patents and so it is worth using TM.

Participant 1

I now receive a lot of PDF files. Can TM products handle them?

Rob Giles

SDLX has a clipboard translator, which allows you to copy and paste the text into the application. However, it is not always possible to copy and paste. Then the only recourse is to go back to the client and ask for the file from which the PDF was generated.

Bjorn Harris

I think the best approach is to print the PDF document, OCR, and export to MS Word.

Michael Benis

You should specify the electronic format with the client. However, if you have to use a PDF file you could get a free download of Acrobat Reader, which will allow you to copy to the clipboard. This will give you a lot of line breaks, which you can get rid of by searching for the breaks and replacing them with a space. If the document is in columns that will mean there is a lot more cleaning up to do and you should agree a fee with the client for this additional work. There are also some packages now which will work with PDF and do the clean-up automatically. They have varying degrees of success. Some OCR packages will also work with PDF. You can specify that you do not want the images, which gives you a smaller file. Generally you do not need the images, but if you do, then agree a rate for handling the translation of these with the client.

Recently OmniPage have produced a plug-in that allows you to convert PDF files to MS Word. I also believe that the increasing use of PDF means that the TM companies are actively seeking ways to handle these.

Barbara Walentynowicz

I am looking for introductory training on how to use TM programmes.

Marc Prior

The manuals are very complex, because the subject is complex. The problem is that there do not seem to be any basic tutorials.

Daniela Ford

This year Imperial College London ran a modular course for ITI members, which I am sure will be repeated. Each weekly workshop covers one TM product, and over an eight to 10 week period, you can attend as many or as few as you wish. The cost is £30 for a two-hour workshop.

Tony Hartley

Before I talk about the eCoLoRe project, we are discussing a course, similar to that run by Imperial College, at Leeds University. This will be a short introduction to TM and probably a weekend course on the products.

The eCoLoRe project is designed to familiarise people worldwide with TM but it is not a course. It is a set of building blocks, in at least 16 languages, designed to allow different countries and groups to tailor TM training to their specific environments. It will include a variety of formats, explanatory information about TM and materials for trainers.

We are one year into the project with 18 months to run. We now have materials ready for release and are looking for people to trial them. Their feedback will then be used to make improvements before we go live to the general public on our website, where contributions will be available as shareware.

Rebecca Power

Could Leeds University incorporate OCR training and how to deal with PDF files into their course?

Tony Hartley

We have the equipment. It is a question of how much we can fit into the course, but the designers will take this suggestion on board.

Michael Benis

More training is being provided now by both the academic community and by the TM manufacturers themselves, who are now collaborating more closely with the universities.

Helen Robertson

Customers increasingly ask us to work with the TM system they use, which is attractive because there is no need to buy software. Is this likely to become more common and will people be disadvantaged if they cannot use TM? Will the systems all be based on one standard and will they be easy to use?

Daniela Ford

I think it is all going to be web-based and because you have to be online all the time you need broadband. Systems that I have used are very easy but can be slow, which is a hardware issue.

Bob Clark

Yes, this will be a big trend with major localisers. It is very flexible and translators can be located anywhere, using a TM interface that is familiar to them, but sharing the same memory. However, it will not replace everything else and it is technology-dependent.

Hayley Harris

How is this going to affect a freelancer's negotiating power?

Michael Benis

TM is attractive to big companies. They can coordinate the work of a large team of translators effectively and get the final result to look as consistent as possible.

Negotiating power depends on how well established you are; what field you are working in; how long you have worked for that client; how desperate you are and to what extent the value-added skills of an individual translator are important to that particular project.

Server-based technology will be used more and more, to get the best possible results in the shortest time. With this in mind it may make more sense to charge services on a time, rather than a word count basis.

At the moment translators gain most from TM in aspects that have nothing to do with translation. This debate and the furore in recent years about differential rates have helped this profession re-assess its value and how it is using TM.

Meike Beumers

Is it fair that clients do not want to pay for 100% matches?

Bob Clark

The answer is better client education. At the moment they do not seem to care when they end up with patchwork results, e.g. style mismatches, etc. The ramifications go beyond TM.

Gerald Dennet

There are genuine reasons for asking translators not to touch 100% matches, e.g. liability issues when lawyers have approved certain words in the safety text. Translators should always be used to check 100% matches, but the client can still choose to ignore this advice.

Participant 2

The problem with partly translated documents is that I may wish to use different terminology in the elements I am translating. This causes me to spend more time on the job than I would do otherwise.

Bjorn Harris

In the software industry there is pressure to get products out on time. This means that if there are problems in the documentation they will deal with these afterwards, rather than compromise the product release schedule.

Marc Prior

Translators are often asked to do things that are inconsistent with their professional ethics. If the translator is asked to do elements of a text, are they being asked to produce a cohesive whole or not? These questions need to be addressed.

Michael Benis

When you are being asked not to touch 100% matches, you still have to read them and translators should be paid for the time involved. The problem is that the accountants have no idea how translation works. A text may have perfect terminological consistency, but it may not make any sense to the target audience. This may impact the perception of product quality because people cannot use it. That will have both an economic impact and a knock-on effect on brand image. The individuals responsible for such practices have not thought through the question of how communications work and the potential impact of poor documentation.

Iwan Davies

The main objective of TM should be to improve the quality of text. Speed should be a side benefit.

Graham Cross

I have had clients who try to help by running a text through TM before sending to me. The 100% matches were correct in the context in which they were originally translated, but the context is now lost. That is a particular problem in those languages, which have limited vocabularies. A word may have multiple meanings and the meaning is defined by the context. I would like to know therefore whether TM works equally well in all languages.

Marc Prior

TM can lull you into a false sense of security. It is dangerous to trust 100% matches.

Gerald Dennet

Part of the problem is the way the various TM products work. If there is a facility to select 'only useful matches' then you will get a better result. However, we cannot train TM to be 100% accurate every time.

Catriona Lischka

What experience does everyone have of the discounts expected by agencies? I have found practices vary.

Charlotte Couchman

I charge on a time basis where there are a lot of 100% matches. However, you must be cautious because sometimes you will receive a better rate on a word count basis. You must be able to dictate the hourly rate if this is the way you are going to work.

Bjorn Harris

Trados have imposed de facto standards on the industry. Perhaps this is an area where the Localisation Industry Standards Association (LISA) could take a role.

Marc Prior

Translators, not the industry and certainly not the TM vendors, should be determining rates.

Gerald Dennet

We pay 33% for perfect matches. Sometimes we tell translators not to touch perfect matches and then we do not pay for them. The reason is that we are reviewing translations in-house to achieve consistency.

Rebecca Power

When you are working for an agency you assume that they are reviewing translations. It is reasonable therefore to give them a discount for taking over this element of the job. This is not the case with end-user clients, who require a fully finished document.

Bob Clark

I would like to know more about the calculations that underlie the banded pricing schemes. Effectively these schemes are cutting into the margins of individual translators who still have to do the work. The wildly different rates suggest that the universities should be brought in to do some research into this and perhaps the well-funded vendors, such as Trados and Microsoft, should sponsor it.

Daniela Ford

Perhaps the ITI in conjunction with the universities should do a rates survey and also get in touch with LISA, with the object of producing guidelines rather than actual rates.

Edmund Jephcott

I am not sure that there is any financial benefit to me in using TM. I believe that the way that it is being used is likely to push a freelance translator's income down.

Bob Clark

Before the advent of TM, translators could make good money on the repetitive jobs and this offset the lower margins on the more difficult work. The lack of collective action in the translation community means clients can dictate terms. However, one part of the solution is not to become a slave to the technology. If you get a fuzzy match, do not agonise over whether to use it. Use the term that you think is best and disregard the machine.

Also, in the early TM products there were no levels of fuzziness. There was a fairly exact match, at a 90% or perhaps 87% level of accuracy, or there was none at all. As a result there were no pricing issues. I believe the time has come to re-examine the whole question of 100% matches, perhaps through a special interest group in LISA.

Gerald Dennet

Increasingly we are being supplied with material from document and content management systems, which our clients increasingly use. As a result we do not get whole texts, we get the bits that have changed. That is the way translation is moving and the customer cannot understand why translators want to charge for work that has already been done. The reality is that we have to give discounts for 100% matches.

Michael Benis

It will be interesting to see whether the texts generated through document and content management systems are effective communications. It may not be economic to use TM for these text fragments and there may very well be a reduction in job satisfaction.

It has been assumed, and the eCoLoRe project is an example, that the reason TM is not used more widely is because there is a shortage of trained users. I believe that translators may not be taking up TM because it is not an economically sound decision for them and also because it decreases job satisfaction. Translators are leaving the industry because their skills mean that they can earn more in other sectors. Consequently there is now a shortage of good translators. Eventually the accountants and the TM vendors may have to start questioning the systems that they use, if they cannot find enough people to use them, on the terms they are offering.

Sarah Slattery

Why are translators interested in tools that are not Microsoft-based and what benefits, if any, are there in using them? What are the advantages of using one tool rather than another?

Marc Prior

It is often more difficult to use alternatives. The biggest single issue is that the tools I use are not 100% compatible with the Microsoft-based tools that everyone uses. The differences are tiny, however, and my customers choose a good translator rather than someone who uses a particular tool. Customers only care about the quality of the ultimate job.

Tools vary in the way that they work, but they are still just tools – some work by sentences, others by paragraphs. You should be looking at the whole text and using the tool as a means to an end.

Daniela Ford

One of the aspects of Linux that attracts users is the perception that it is a more secure system. I take the easy route and use the same system that my customers use, so I can be sure that their job is in the right format.

Participant 3

Is the choice of tool text-dependent?

Michael Benis

TM products are very similar. It is mainly a matter of personal preference, but using the same system a client uses can save a lot of work and worry.

Daniela Ford

Many agencies supply software free, so you do not have to pay for lots of different systems.

Tony Hartley

Translation Memory Exchange (TMX), pioneered by Open Standards for Container/Content Allowing Re-use (OSCAR), is intended to provide greater openness so that different systems can share information. In theory this is good news. It allows you to choose the TM application with which you are most comfortable. However, it would be interesting to know how much emphasis vendors are putting on TMX when they develop their products.

Gerald Dennet

The latest version of Transit is TMX-compliant for import and export of text, although we have never been asked to supply or receive TMX material. The product is currently going through the certification process.

Bjorn Harris

TMX1 related simply to the import and export of text. TMX2 exports the text and the formatting.

Rob Giles

Early in TMX development there were some problems because of the different ways in which vendors interpreted the TMX specification. This meant that you could produce a TM text from one tool and you would not be able to import it into another. Recently the OSCAR steering committee has closed those loopholes and there is now a proper certification process. So, if you are looking at TMX, make sure that they are TMX-certified products. We are taking TMX very seriously. SDLX has achieved certification.

Michael Benis

TMX is useful for people like me who test many different software packages. They convert the database or its equivalent from one format to another, so that you can use different products but use your own memory all the time. Similarly if you are sent a memory, which a client wants you to use, you can convert it to work with your own system. It is also useful for people who are migrating from one system to another.

You will still have to convert the text that you are sent if you are using a different product to the one your client is using. That is likely to be your major concern as a translator, rather than converting the databases.

Bob Clark

If you are going to use the eCoLoRe site, the materials that reside on it will almost certainly be in TMX format. Whether it works or not remains to be seen.

ITI TM debate - Panel biographies

Michael Benis studied at Christ's College, Finchley, and the University of York, where he graduated in English and Related Literature, followed by research and teaching. He has been a full-time translator and interpreter since 1984 and is a member of the Institute of Translation and Interpreting. He currently combines translating and interpreting with copywriting, speech writing and journalism. He has published widely on the integration of new technologies into the translation process.

Bob Clark followed a lengthy spell as a linguist with the US Defence Department by settling in Yorkshire and took a Joint Honours Degree in Arabic and Russian with Bulgarian at University of Leeds. Studied computer programming and systems analysis and design. Following a spell running his own business his re-entry to the Translation Industry coincided with the embryonic development of translation tools. He was appointed Software Editor of *Language International* magazine and eventually went on to be Co-Editor with Bert Esselink. Currently he holds the post of Senior Lecturer in Translation Technology at the University of Leeds, he works as a Solutions Consultant with the Logos Group of Companies in Modena, Italy, and occasionally sits in as Guest Editor of the *LISA Globalization Insider* newsletter.

Charlotte Couchman a founder member of ITI, graduated in French and German from Bristol University in 1984. She took a postgraduate diploma in translation at University of Kent, Canterbury and has worked freelance ever since. Most of her work has been patents, and together with husband Paul Clarke set up a partnership, Lodestar Translations, a few years ago and branched out into other, generally more interesting areas, such as environmental reports and Websites.

Daniela Ford, MITI, has an MSc in Technical Translation from Hildesheim University in Germany. She has been a staff translator for 5 years and has been working as a freelance translator for 3 years. Her main areas of translation are software localisation and technical manuals, involving the use of translation memory software on a daily basis. She is also a part-time lecturer on the MSc Translation course at Imperial College London where her specialist subjects include translation memory technology. Daniela is the current Chairman of the London Regional Group of ITI and Member of the ITI Council. She gives regular talks in the UK on software localisation, translation technology and web site design.

Rob Giles has worked for SDL International for over 3 years, gaining a great deal of knowledge and expertise in SDLX Translation Suite - SDL's Computer Aided Translation tool. He started his SDL career as one of the support engineers for SDL's range of software products. He then moved to the Desktop Products Division where he now deals with pre-sales support for the U.S. market, and hosts both online and onsite training for SDLX.

Bjorn Harris is a Sales and Support manager for ATRIL UK, who sell and market the Translation Memory system Déjà Vu in the UK and Ireland. Bjorn has a degree in Business Information Management from Leeds Metropolitan University and he has extensive knowledge of the translation and localisation industry gained over 5 years. Initially starting out at ALPNET UK managing the workflow process of its proprietary TM tool - Joust, Bjorn became involved in Project Management before joining the Déjà Vu team. Bjorn is currently responsible for the sales and support in the UK and Ireland of the entire new DVX product range which includes DVX Workgroup, DVX TM Server, DVX TM Builder, DVX Professional, DVX Standard and DVX Editor.

Tony Hartley is Professor of Translation Studies and Director of the Centre for Translation Studies, University of Leeds. In his early academic career, he worked extensively as a conference interpreter, before moving into the field of human language technologies. He has been involved in UK and EU sponsored research into the computer generation of multilingual documents and in the EUROMAP initiative. His current work focuses on the evaluation of machine translation output quality. He is co-ordinator of the eCoLoRe project.

Marc Prior studied German and Italian at Bangor (Wales) and translation at Kent (Canterbury). A staff translator's job in the electrical engineering industry was followed in 1991 by a move to Germany (Cologne), and two years later by the decision to become self-employed. His specialist translation fields are now occupational health and safety, and environmental engineering. In 1999 he discovered the Linux operating system, and since then has invested - or wasted, depending upon your point of view - many hours exploring and documenting its suitability for use by translators. He has also acquired responsibility for the free OmegaT translation memory program.