

GUIDANCE ON TECHNOLOGY FOR TRANSLATORS

Introduction

Translation is an act of writing that conveys the meaning of a text written in one language in another language. The history of translation keeps pace with the story of writing and the discovery of the science and literature of other cultures. Evidence exists of translations of a Sumerian epic dating from the second millennium BC. A journey through the centuries reveals significant translation activities in Ancient China, Mesopotamia, Rome, Toledo and the court of King Alfred the Great. Translation brought knowledge from the Orient to Europe, and the translations of the Bible into German and English were instrumental in changing the course of history. For all the potential for impact of his or her craft, however, the translator's essential tools were and still are simple ones: an instrument for writing and a medium on which to write. All a translator fundamentally needs to render the Gospel of Saint Mark, a challenging essay by Heidegger or a contemporary Japanese novel into a comprehensible piece of Welsh (or English) is a fine mind, flair, an A4 pad and a pencil.

The second half of the twentieth century saw the professionalisation of translation. Institutes and organisations were founded, standards and qualifications were developed, and universities introduced courses designed to prepare their graduates for admission to this new profession. Newly qualified translators will start their careers as freelancers working for Language Service Providers (LSP's) or as staff translators for a government or commercial organisation. Some will eventually set up their own businesses, while others will remain in employment throughout their working lives. The growing demand for translation world-wide has focused commercial and government organisations on the cost of translation and the need to maximise translator productivity. Translation technology has been identified as a means of boosting productivity. The one constant in the varied career paths of professional translators is that they will all be expected to be proficient in the use of an array of translation tools.

Technology for translators

Over the past thirty years translators have benefited from the development of a range of computer assisted translation tools and other technology. These tools include:

- Translation memory software
- Machine translation software
- Electronic dictionaries and online glossaries
- Voice recognition software
- OCR and PDF conversion software
- Spelling and grammar checkers

This classification is simplistic insofar as most translation toolsets nowadays combine two or more of these tools within a single package. However, it is useful to understand the function of each component before looking at ways they can be combined in a practical translation environment.

Translation memory software

Translation memory software is likely to be the tool translators will use most in their careers. The underlying principle of this software, commonly known as TM software, is that a translator should never have to translate the same sentence twice. In its most basic form a translation memory application stores each translated sentence in a database which it then consults to translate subsequent sentences. If the source sentence has a translation in the database, there is a 100% match, the reliability of which is dependent on the underlying reliability of the translation memory. The translator may also be presented with partial matches (“fuzzy matches”) which can be accepted, edited or rejected. This is the basic function of the translation memory program, but the most advanced TM packages go far beyond that and incorporate sophisticated terminology management, workflow and quality assurance services.

In theory, translators can start with an empty database, or translation memory, and build up their own memory sentence by sentence. Over several years, the professional translator will build up a substantial database of translated sentences, which will save a significant amount of time and effort, particularly if he or she only undertakes work in a specialised field. Translators will also receive existing translation memories from clients or make use of freely available public translation memories. The European institutions and organisations such as TAUS provide opportunities for sharing data in the form of translation memories and their databases can be consulted free of charge. The TMX and XLIFF file formats have been developed to facilitate the exchange of such language resources.

The principal commercial translation memory programs used by translators today are SDL Trados Studio 2014, memoQ 2014, Wordfast Pro and Déjà Vu X3. However, there are at least 25 translation memory tools available, including a number of open source programs such as Omega-T and OpenTM2. Several of these programs started life as macros or templates to be run within Microsoft Word. Nowadays, apart from the legacy Trados Translator's Workbench and Wordfast Classic, all the translation memory packages run as self-contained programs within some kind of graphical user interface (GUI) or within a browser.

Modern commercial CAT tools like SDL Trados Studio 2014 can handle over 50 different file formats. This enables translators to translate text in applications such as Adobe InDesign and Microsoft PowerPoint without needing to have those programs installed on their PC. As the software will support any language combination that is supported by Microsoft Windows, the translator can translate between languages with

completely different scripts, e.g. from Bengali into Russian.

The user typically works within a graphical interface with the source and target languages displayed alongside each other. For each new segment (usually a sentence) in the source text, proposed translations are displayed in a separate pane, allowing the translator to select the most appropriate translation which is then inserted in the editor window alongside the source sentence. The translator may also decide to use one or more machine translation plug-ins which will come up with translation proposals if there is no corresponding sentence in the selected translation memory. The pros and cons of using this kind of plug-in will be discussed below.

Translation memory software is designed for interactive use. However, once the translator has built up or obtained a large enough translation memory he or she will achieve significant time savings by automating part of the translation process with the software's "pre-translate" feature. This feature will attempt to translate the document by first retrieving 100% matches or fuzzy matches from the selected translation memories and then by using a machine translation plug-in if this option has been chosen by the user. At the end of this automated process it is considered good practice for the translator to review the generated translation and approve or edit each segment in turn.

Freelance translators usually work with a translation memory installed on their own PCs, even if it is provided by a client. Staff translators or freelance translators working on large projects involving several translators may make use of a server-based translation memory. In such circumstances translators are able to share the benefits of each other's translation work and are more likely to make consistent use of terminology.

The concept of the server-based translation environment has over the last couple of years undergone a progression into what is known as the cloud-based translation environment. Memsources, XTM, SmartCAT, MateCAT, Wordfast Anywhere, Smartling, Easyling, Lingotek, and Wordbee are completely web-based solutions. While some of these solutions are fee or subscription based, MateCAT is free open source software released under the LGPL license.

MateCat ((Machine Translation Enhanced Computer Assisted Translation) has been developed with public funds by some of Europe's leading researchers in the field of language technology and has been designed with ease of use in mind. The translator can quite literally enter <https://www.matecat.com/> in his or her browser and start using the software, which supports 56 file formats and languages extending from Afrikaans to Welsh, including languages with non-Latin alphabets and right-to-left languages. User support is provided free of charge via a Facebook group and via e-mail. As its name suggests MateCAT combines Machine Translation and Translation Memory technologies. It automatically links to the freely available TM Server My Memory and to Google Translate and will therefore come up with translation

suggestions for the Welsh-English/English-Welsh language pair.

Machine translation

Machine translation, also known as automatic translation or computer-generated translation, is the branch of language technology in which a computer program translates without any human intervention. The usefulness of these translations is dependent on a variety of factors including language pair, the complexity or ambiguity of the source text and the familiarity of the system with the terminology and subject matter of the original. Some languages are more amenable than others to machine translation. Welsh, with its mutations and sentence structure, poses a significant though not insurmountable challenge to developers.

Research and development in machine translation had its first big boost in the Cold War period with much attention being paid to Russian and Chinese. The expansion of the European institutions brought about investment in the development of the commercial Systran software to cover the languages used at the European Commission. International conflicts and the need for better intelligence have also given an impetus to the development of machine translation programs for languages such as Arabic, Pashto and Farsi, particularly in the USA. But the need to translate many of the 684,478 pieces of content generated every minute on the Internet is perhaps the most powerful driving force behind progress in MT R&D. In Wales the introduction of machine translation by the National Assembly for Wales has to be seen in the context of the Official Languages Scheme. It is a tool to help people living and working in Wales to communicate in either of the official languages.

Much of the earlier work in MT produced Rule Based and Example Based systems. Since the mid-1990's Statistical Machine Translation, where translations are generated on the basis of statistical models derived from the analysis of bilingual text corpora has become the predominantly used approach, although researchers in this field are now working on the introduction of syntactic rules into the translation process. Both publicly available machine translation systems offering Welsh-English/English-Welsh – Google Translate and Microsoft Translator – are fundamentally statistical systems. The successful use of these systems is broadly dependent on the quantity and quality of the bilingual data used to build the translation model. The partnership between the National Assembly for Wales and Microsoft which provides a mechanism for submitting a variety of bilingual texts to the system is the best guarantee for achieving a steady improvement in the quality of the computer-generated translations.

The general public can access Microsoft Translator directly via a browser under any operating system (including Linux and Android) or from within the Microsoft Office suite where there is greater functionality than when simply using the browser. An employee can receive an e-mail in Welsh and read an English translation at the click of a mouse. Professional translators may prefer to use the Welsh-English/English-

Welsh MT service as a plug-in within their chosen CAT environment. Most translation memory applications offer machine translation plug-ins and allow the translator to choose when a MT-based translation proposal will be made. The translator can also “pre-translate” an entire document so that Microsoft Translator will populate all the segments for which there is not a translation in the translation memory. Whichever approach is used, the translator is able to review or post-edit the machine translation proposal so that only an approved translation is saved in the translation memory. Studies have shown that significant productivity gains can be made by combining machine translation with post-editing in a translation environment tool.

Terminology tools

Translators who don't use machine translation and translation memory software still have a range of tools available that will help them to use the correct terminology in their translations. Many of these are online and available free of charge. The Welsh Academy Dictionary Online (<http://www.geiriaduracademi.org>) and the Welsh National Terminology Portal (www.termaw.org) are world-class lexicographical resources which can be consulted through any browser. The multilingual database MyMemory (www.mymemory.translated.net) contains numerous Welsh-English entries which are likely to be drawn from the Corpus of the Welsh Assembly Record of Proceedings (www.corpws.org/ycofnod) which can be searched directly using the Cysefin and Hebog search tools developed by the Language Technology Unit of Bangor University.

Spelling and grammar checkers

Proofing tools such as the Cysill Welsh online spelling and grammar checker (<http://www.cysill.com/arlein>), the spellchecker in Microsoft Word, E-gyhoeddwr (E-publisher) are quality assurance tools that enable a translator to deliver professional work without embarrassing spelling or grammatical mistakes. The site giving Welsh equivalents of English place names in Wales (<http://www.e-gymraeg.org/enwaucymru/>) will provide translations of place names that do not occur frequently enough in official documents to make their way into the bilingual corpora. These tools are useful quality safeguards for those translation jobs where translation automation is impossible, inappropriate or offers no productivity gain.

Voice recognition software

Voice recognition is not yet available for translators working into Welsh, but there are several voice recognition packages on the market which will enable those working into English to achieve significant productivity gains. Nuance's *Dragon Naturally Speaking* range is the market leader in speech recognition. The *Home* version is an entry-level product that is adequate for occasional use; translators who want to make voice recognition their principal tool for producing their translations would probably

benefit from upgrading to the professional version. The big advantage of voice recognition is that it integrates fully with any other tool running under one of the recent versions of Microsoft Windows. This means that the translator can use voice recognition within a translation memory package, even to post-edit machine translation output.

OCR & PDF conversion software

Optical Character Recognition (OCR) software is an essential tool in the translation kit for those translators who want to use CAT tools but receive their jobs in the form of hard copy. Nowadays most printers are multifunctional machines that can print, scan and copy so it is not necessary to purchase dedicated hardware. This software allows the translator to scan the text to be translated and save it in an editable format (e.g. as a Word or Excel file). Two decades ago such software cost thousands of pounds, though today there are many free packages available on the Internet. However, it is important to use OCR software with a proven level of accuracy on medium-quality documents, otherwise the productivity gain achieved by not having to type the text will be lost through the need to correct pages of “OCR garbage”. Nuance's entry-level Omnipage 18 and the Abbyy FineReader are considered to be a reliable applications for turning printed documents into electronic files that can be processed by machine translation and/or translation memory software.

Many translators receive documents as PDF files which they will have to convert into editable text if they intend to use CAT tools. Some of these CAT tools claim to handle PDF files as one of their file types, but they will be using a third-party PDF converter “behind the scenes” and it is unlikely that the translator will be able to influence the conversion process. Dedicated PDF conversion programs such as Nuance's PDF Converter or the Abbyy PDF Transformer will generally provide results that need little or no manual correction.

Many “free” OCR and PDF conversion packages are available on the Internet. Translators should exercise caution before downloading such packages. Many of them are inadequately supported and may also serve to install malware onto the user's PC.

Translators and security

All professional translators have a duty of care towards their clients. Translators are expected to deliver accurate, appropriately presented translations, but also have a duty of confidentiality. This duty needs to be reflected in the level of security translators implement in their work environment. Simple precautions such as locking up confidential documents and not leaving unprotected computers switched on unattended need to be complemented by effective cybersecurity measures. Judicious use of public WiFi, encryption of sensitive documents sent via e-mail, precautions against third-party use of USB-sticks and smart phones in the vicinity of computers

holding confidential or sensitive information, effective removal of files from hard drives are common-sense measures which are often ignored.

Translators working throughout the United Kingdom have to comply with the Data Protection Act 1998 if their work involves documents containing personal information on third parties. They are not permitted to hold this data longer than is strictly necessary for their work and must keep it safe and secure. Many translators back up their data to cloud-based storage systems such as Google Drive and Dropbox. Both these services claim to apply the highest security standards and offer safeguards such as encryption and login authentication. However, it is good practice for the translator to consult his/her clients before backing up their data to the cloud, and the servers on which their data is stored must always be located within the European Union.

Similar considerations apply to the use of cloud-based machine translation and translation memory systems. Google clearly states that it stores all the information submitted to its translation servers. SDL states that content submitted to the SDL Language Cloud API is not stored on the server after a translation is completed. The Microsoft Translator service is governed by the Microsoft Service Agreement and the Microsoft Online Privacy Statement (<http://msdn.microsoft.com/en-us/library/hh464486.aspx>). Translators should read these documents and decide whether they can meet applicable confidentiality requirements before they submit documents to the Microsoft Translator servers.

Translator training

Younger translators will probably have acquired some basic IT skills at home and at school. Translators who are not comfortable with computers will find the use of translation tools a challenge. A good step for them would be to take some basic IT courses. 'Computer Courses Near You' is a Welsh Government/European funded online resource that is used by staff in libraries and post offices to signpost enquiries to suitable IT courses. Customers looking for computer courses can go to any Post Office counter and ask where their nearest computer courses are held. They will receive a print out showing the nearest venues to them. This information is also the primary resource for librarians who deal with a large volume of enquiries about computer courses because of the free internet access that is now offered in most libraries.

Cymdeithas Cyfieithwyr Cymru (The Association of Welsh Translators and Interpreters) and the Institute of Translation and Interpreting (ITI) organise reasonably-priced workshops and webinars on the use of translation tools as part of their CPD programmes. Major vendors of translation tools such as SDL (SDL Trados Studio 2014) and Kilgray Translation Technologies (memoQ) provide online training in their products and face-to-face training through their accredited trainers. Most vendors provide some form of support through forums and chat rooms, and details of

many of the available training courses and events are given under the Education tab of the website of proz.com – an online marketplace for translators (www.proz.com/about/overview/education). Users of translation tools will testify that taking some basic training is the most effective way of getting up and running with a new CAT package.

Further reading

- Jost Zetzsche's *The ToolBox Journal* (described by the author as “A monthly journal for people in the translation industry who want to get more out of their computers”). For details see <http://www.internationalwriters.com/toolkit>.
- *101 Things a Translator Needs to Know* - a collection of tips from a group of experienced translators. For details see <http://101things4translators.com>.
- *Machine Translation*, W.H. Hutchins & H.L.Somers, Academic Press Inc, 1992. This is the first textbook of machine translation, providing a full course on both general machine translation systems characteristics and the computational linguistic foundations of the field. It does not discuss Statistical Machine Translation.
- Statistical Machine Translation Tutorial Reading (<http://www.cs.pomona.edu/~dkauchak/mt-tutorial/>): a series of links to courses and papers for those who want to know more about this field.
- *PCs for Dummies*, Dan Gookin, John Wiley & Sons; 9th Edition edition (29 Sept. 2003) . A good starting point for those taking their first steps in computing.

**Terence Lewis MITI
Ludlow, January 2015**