Introduction

When the term computer-aided translation is mentioned, we often associate it with the functions a computer-aided translation system offers, such as toolbars, icons, and hotkeys, the built-in tools we can use, such as online dictionaries, browsers, and the computational hitches we often encounter when working on a computer-aided translation project, such as chaotic codes. What is more important is to see beyond the surface of computer-aided translation to find out the major concepts that shape the development of functions in translation technology.

Concepts, which are relatively stable, govern or affect the way functions are designed and developed, while functions, which are fast-changing, realize the concepts through the tasks they perform. As a major goal of machine translation is to help human translators, a number of functions in computer-aided translation systems have been created to enable machine processing of the source with minimum human intervention. Concepts, moreover, are related to what translators want to achieve in translating. Simply put, translators want to have a controllable (controllability) and customizable (customizability) system, which is compatible with file formats (compatibility) and language requirements, and behaves as well as (simulativity) or even better than (emulativity) a human translator, to allow them to work together (collaborativity) to produce quality translations (productivity). We have therefore identified seven major concepts which are important in computer-aided translation: simulativity, emulativity, productivity, compatibility, controllability, customizability, and collaborativity. The order in which concepts are arranged can be memorized more easily by their acronym SEPCCCC.

Simulativity

The first concept of computer-aided translation is simulativity, which is about the way in which a computer-aided translation system models the behaviour of a human translator by means of its functions, such as the use of concordancers in text analysis to model after comprehension on the part of the human translator and the creation of a number of quality assurance tools to follow the way checking is done by a human translator.

There are a number of ways to illustrate man–machine simulativity.
(1) Goal of translation

The first is about the ultimate goal of translation technology. All forms of translation (machine translation, computer-aided translation and human translation) aim at obtaining high-quality translations. In the case of machine translation, the goal of a fully automatic high-quality translation (FAHQT) is to be achieved through the use of a machine translation system without human intervention. In the case of computer-aided translation, the same goal is to be achieved through a computer-aided translation system that simulates the behaviour of a human translator through man–machine interaction.

(2) Translation procedure

A comparison of the procedures of human translation with those of computer-aided translation shows that the latter simulates the former in a number of ways. In manual translation, various translation procedures have been proposed by translation scholars and practitioners, ranging from the two-stage models to eight-stage ones, depending on the text type and purposes of translation. In machine translation and computer-aided translation, the process is known as technology-oriented translation procedure.

(a) Two-stage model

In human translation, the first type of translation procedure is a two-stage one, which includes the stage of source text comprehension and the stage of target text formulation, as shown below:

Figure 2.1  A two-stage model for human translation

Figure 2.1 is a model for human translators with the ability of comprehension. As a computer-aided translation system does not have the ability of comprehension, it cannot model after human translation with this two-stage model. It can, however, work on a two-stage translation with the use of its system dictionary, particularly in the case of a language-pair-specific system, as in Figure 2.2:

Figure 2.2  A two-stage dictionary-based language-pair-specific model
Another two-stage model of computer-aided translation is a terminology-based system, as shown in Figure 2.3:

![Figure 2.3](image)

Figure 2.3 A two-stage terminology-based CAT system

(b) Three-stage models

The second type of translation procedure is a three-stage model. This section covers five variations of this model proposed by Eugene Nida and Charles Taber (1969), Wolfram Wils (1982), Roger Bell (1991), Basil Hatim and Ian Mason, and Jean Delisle (1988) respectively. A three-stage example-based computer-aided translation system is shown to illustrate the simulation of human translation by computer-aided translation.

(i) MODEL BY EUGENE NIDA AND CHARLES TABER

The first model of a three-stage translation procedure involving the three phases of analysis, transfer, and restructuring was proposed by Eugene Nida and Charles Taber ([1969] 1982: 104). They intended to apply elements of Chomsky’s transformational grammar to provide Bible translators with some guidelines when they translate ancient source texts into modern target texts, which are drastically different in languages and structures. Nida and Taber describe this three-stage model as a translation procedure in which

the translator first analyses the message of the source language into its simplest and structurally clearest forms, transfers it at this level, and then restructures it to the level in the receptor language which is most appropriate for the audience which he intends to reach.

(Nida and Taber [1969] 1982: 484)

Analysis is described by these two scholars as ‘the set of procedures, including back transformation and componential analysis, which aim at discovering the kernels underlying the source text and the clearest understanding of the meaning, in preparation for the transfer’ (Nida and Taber [1969] 1982: 197). Transfer, on the other hand, is described as the second stage ‘in which the analysed material is transferred in the mind of the translator from language A to language B’ (ibid.: 104). Restructuring is the final stage in which the results of the transfer process are transformed into a ‘stylistic form appropriate to the receptor language and to the intended receptors’. 
In short, analysis, the first stage, is to analyse the source text, transfer, the second stage, is to transfer the meaning, and restructuring, the final stage, is to produce the target text. Their model is shown in Figure 2.4.

(iii) MODEL BY ROGER BELL

Another three-stage model of note is by Roger Bell whose translation procedure framework is divided into three phases: the first phase is source text interpretation and analysis, the second, translation process, and the third, text reformulation (see Figure 2.6). The last phase takes into consideration three factors: writer’s intention, reader’s expectation, and the target language norms (Bell 1991).
This model, proposed by Basil Hatim and Ian Mason, is a more sophisticated three-stage model, which involves the three steps of source text comprehension, transfer of meaning, and target text assessment (see Figure 2.7). At the source text comprehension level, text parsing, specialized knowledge, and intended meaning are examined. At the meaning transfer stage, consideration is given to the lexical meaning, grammatical meaning, and rhetorical meaning. At the target text assessment level, attention is paid to text readability, target language conventions, and the adequacy of purpose.

The fourth model of a three-stage translation procedure was proposed by Jean Delisle (1988: 53–69) (see Figure 2.8). Deslisle believes that there are three stages in the development of a translation equivalence: comprehension, reformulation, and verification: ‘comprehension is
Computer-aided translation

based on decoding linguistic signs and grasping meaning, reformulation is a matter of reasoning by analogy and re-wording concepts, and verification involves back-interpreting and choosing a solution’ (1988: 53).

Parallel to human translation, a three-stage model in computer-aided translation is the example-based system. The input text goes through the translation memory database and glossary database to generate fuzzy matches and translations of terms before getting the target text. The procedure of an example-based computer-aided translation system is shown in Figure 2.9.

![Three-stage example-based computer-aided translation model](image)

Figure 2.9  Three-stage example-based computer-aided translation model

(c) Four-stage model

The third type of translation procedure is a four-stage one. A typical example is given by George Steiner ([1975] 1992) who believes that the four stages of translation procedure are: knowledge of the author’s times, familiarization with author’s sphere of sensibility, original text decoding, and target text encoding (see Figure 2.10).

![Four-stage model](image)

Figure 2.10  Four-stage model of translation
For computer-aided translation, a four-stage model is exemplified by webpage translation provided by Yaxin. The first stage is to input the Chinese webpage, the second stage is to process the webpage with the multilingual maintenance platform, the third stage is to process it with the terminology database, and the final stage is to generate a bilingual webpage. The Yaxin translation procedure is shown in Figure 2.11.

(d) Five-stage model

The fourth type of translation procedure is a five-stage one, as proposed by Omar Sheikh Al-Shabab (1996: 52) (see Figure 2.12). The first stage is to edit the source text, the second, interpret the source text, the third, interpret it in a new language, the fourth, formulate the translated text, and the fifth, edit the formulation.

In computer-aided translation, a five-stage model is normally practised. At the first stage, the Initiating Stage, tasks such as setting computer specifications, logging in a system, creating a profile, and creating a project file are performed. At the second stage, the Data Preparation Stage, the tasks involve data collection, data creation, and the creation of terminology and translation memory databases. At the third stage, the Data Processing Stage, the tasks include data analysis, the use of system and non-system dictionaries, the use of concordancers, doing
Computer-aided translation

pre-translation, data processing by translation by computer-aided translation systems with human intervention, or by machine translation systems without human intervention, or data processing by localization systems. At the fourth stage, the Data Editing Stage, the work is divided into two types. One type is data editing for computer-aided translation systems, which is about interactive editing, the editing environments, matching, and methods used in computer-aided translation. Another type is data editing for computer translation systems, which is about post-editing and the methods used in human translation. At the last or fifth stage, the Finalizing Stage, the work is mainly updating databases.

The five stages in computer-aided translation are illustrated in Figure 2.13.

It can be seen that though there are both five-stage models in human translation and computer-aided translation and the tasks involved are different, the concept of simulativity is at work at almost all stages.

(e) Eight-stage model

The fifth type of translation procedure is an eight-stage one, as proposed by Robert Bly (1983).
Robert Bly, who is a poet, suggests an eight-stage procedure for the translation of poetry: (a) set down a literal version; (b) find out the meaning of the poem; (c) make it sound like English; (d) make it sound like American; (e) catch the mood of the poem; (f) pay attention to sound; (g) ask a native speaker to go over the version; and (h) make a final draft with some adjustments (see Figure 2.14).

Figure 2.14  Model by Robert Bly (1983)

In computer-aided translation, there is no eight-stage model. But other than the five-stage model, there is also a seven-stage model, which is shown in Figure 2.15.

Figure 2.15  Seven-stage computer-aided translation procedure

The seven stages of computer-aided translation go from sample text collection to termbase creation, translation memory database creation, source text selection, data retrieval, source text translation and finally data updating.
All in all, we can say that when compared to human translation, computer-aided translation is simulative, following some of the stages in human translation.

**Emulativity**

There are obviously some functions which are performable by a computer-aided translation system, but not by a human translation. This is how technology can emulate human translation. Computer-aided translation, with the help of machine translation, simulates human translation, and it also emulates human translation in a number of areas of computer-aided translation, some of which are mentioned below.

### Alt-tag translation

This function of machine translation allows the user to understand the meaning of text embedded within images (Joy 2002). The images on a web site are created by IMG tag (inline image graphic tag), and the text that provides an alternative message to viewers who cannot see the graphics is known as ALT-tag, which stands for ‘alternative text’. Adding an appropriate ALT-tag to every image within one’s web site will make a huge difference to its accessibility. As translators, our concern is the translation of the alternative text, as images are not to be translated anyway.

### Chatroom translation

Machine translation has the function to translate the contents of a chatroom, known as ‘chat translation’ or ‘chatroom translation’. Chat translation systems are commercially available for the translation of the contents of the Chatroom on the computer. As a chat is part of conversational discourse, all the theoretical and practical issues relating to conversational discourse can be applied to the study of chat translation. It should be noted that this kind of online jargon and addressivity are drastically different from what we have in other modes of communication.

The function of Chatroom is available in some systems, such as Fluency, as one of the resources. This function has to be purchased and enabled in the Fluency Chat Server to allow clients to be connected to this closed system for internal communications. For standalone version users, the function of Chat will be provided by Fluency Chat Server provided by its company, the Western Standard (Western Standard 2011: 39).

### Clipboard translation

This is to copy a text to the clipboard from any Windows application for a machine translation system to translate the clipboard text, and the translated text can then be pasted in the original or any other location. One of the systems that translate clipboards is Atlas.

### Conversion between metric and British systems

A function that can be easily handled by machine translation but not so easily by human translation is the conversion of weight, volume, length, or temperature from metric to British or vice versa. Fluency, for example, can do the metric/British conversion the target text box with the converted units.
**Currency conversion**

Some computer-aided translation systems can do currency conversion. With the use of Currency Converter, a function in Fluency, and access to the Internet to get the currency conversion rates, systems can convert a currency in a country into the local country currency. The number of currencies that can be handled by a system is relatively large. Fluency, for example, supports the conversion of currencies of around 220 countries. The conversion of multiple currencies is also supported.

**Email translation**

This refers to the translation of emails by a machine translation system (Matsuda and Kumai 1999; Rooke 1985: 105–115). The first online and real-time email translation was made in 1994 by the CompuServe service which provided translation service of emails from and to English and French, German or Spanish. Email translation has since become a very important part of daily communication and most web translation tools have email translators to translate emails. As emails are usually conversational and often written in an informal or even ungrammatical way, they are difficult for mechanical processing (Fais and Ogura 2001; Han, Gates and Levin 2006). One of the systems that translates emails is Atlas.

**Foreign language translation**

One of the most important purposes of using translation software is to translate a source text the language of which is unfamiliar to the user so as to explain its contents in a language familiar to the user. It is found that a majority of the commercial machine translation systems are for translation among Indo-European languages or major languages with a large number of speakers or users. Software for translation between major languages and minor languages are relatively small in number.

**Gist translation**

Another area where machine translation differs fundamentally from human translation is gist translation, which refers to a translation output which expresses only a condensed version of the source text message. This type of rough translation is to get some essential information about what is in the text for a user to decide whether to translate it in full or not to serve some specific purposes.

**Highlight and translate**

This function allows the user to highlight a part of the text and translate it into the designated language. The highlighted text is translated on its own without affecting the rest of the text.

**Instant transliteration**

This refers to a function of machine translation which can transliterate the words of a text with a certain romanization system. In the case of Chinese, the Hanyu Pinyin Romanization system for simplified characters is used in mainland China, while the Wade-Giles Romanization system for traditional characters is used in Taiwan.
Computer-aided translation

Mouse translation
This is to translate sentences on a web page or on applications by simply clicking the mouse. Systems that provide mouse translation include Atlas.

Online translation
This is the translation of a text by an online machine translation system which is available at all times on demand from users. With the use of online translation service, the functions of information assimilation, message dissemination, language communication, translation entertainment, and language learning can be achieved.

Pre-translation
Machine translation is taken to be pre-translation in two respects. The first is as a kind of preparatory work on the texts to be translated, including the checking of spelling, the compilation of dictionaries, and the adjustment of text format. The second is taken to be a draft translation of the source text which can be further revised by a human translator.

Sentence translation
Unlike human translation which works at the textual level, machine translation is sentential translation. In other words, machine translation is a sentence-by-sentence translation. This type of translation facilitates the work of post-editing and methods which are frequently used in translating sentences in translation practice to produce effective translations can be used to produce good translations from machine translation systems.

Web translation
This refers to the translation of information on a web page from one language into another. Web-translation tools are a type of translation tools which translate information on a web page from one language into another. They serve three functions: (1) as an assimilation tool to transmit information to the user; (2) as a dissemination tool to make messages comprehensible; and (3) as a communication tool to enable communication between people with different language backgrounds.

Productivity
As translation technology is a field of entrepreneurial humanities, productivity is of great importance. Productivity in computer-aided translation is achieved through the use of technology, collective translation, recycling translations, reusing translations, professional competence, profit-seeking, labour-saving, and cost-saving.

Using technology to increase productivity
The use of technology to increase productivity needs no explanation. As early as 1980, when Martin Kay discussed the proper place of men and machines in language translation, he said:
Translation is a fine and exacting art, but there is much about it that is mechanical and routine and, if this were given over to a machine, the productivity of the translator would not only be magnified but his work would become more rewarding, more exciting, more human.

(Kay 1980: 1)

All computer-aided translation systems aim to increase translation productivity. In terms of the means of production, all translation nowadays is computer-aided translation as virtually no one could translate without using a computer.

Collective translation to increase productivity

Gone are the days when bilingual competence, pen and paper, and printed dictionaries made a translator. Gone are the days when a single translator did a long translation project all by himself. It is true that in the past, translation was mainly done singly and individually. Translation was also done in a leisurely manner. At present, translation is done largely through team work linked by a server-based computer-aided translation system. In other words, translation is done in a collective manner.

Recycling translations to increase productivity

To recycle a translation in computer-aided translation is to use exact matches automatically extracted from a translation memory database. To increase productivity, the practice of recycling translations is followed in computer-aided translation. Networked computer-aided translation systems are used to store centralized translation data, which are created by and distributed among translators. As this is the case, translators do not have to produce their own translations. They can simply draw from and make use of the translations stored in the bilingual database to form their translation of the source text. Translation is therefore produced by selection.

Reusing translations to increase productivity

To reuse a translation in computer-aided translation is to appropriate terms and expressions stored in a term database and translation memory database. It should be noted that while in literary translation, translators produce translations in a creative manner, translators in practical translation reuse and recycle translations as the original texts are often repetitive. In the present age, over 90 per cent of translation work is in the area of practical translation. Computer-aided translation is ideal for the translation of repetitive practical writings. Translators do not have to translate the sentences they have translated before. The more they translate, the less they have to translate. Computer-aided translation therefore reduces the amount a translator needs to translate by eliminating duplicate work. Some systems, such as Across, allow the user to automatically reuse existing translations from the translation memory. It can be seen that ‘reduce, reuse, recycle’ are the three effective ways of increasing profitability (de Ilarraza, Mayor and Sarasola 2000).
Professional competence to increase productivity

Translators have to work with the help of translation technology. The use of computer-aided translation tools has actually been extended to almost every type of translation work. Computer-aided translation tools are aimed at supporting translators and not at replacing them. They make sure that translation quality is maintained as ‘all output is human input’. As far as the use of tools is concerned, professional translation is technological. In the past, translators used only printed dictionaries and references. Nowadays, translators use electronic concordancers, speech technology, online terminology systems, and automatic checkers. Translation is about the use of a workbench or workstation in translation work.

Translation competence or knowledge and skills in languages are not enough today. It is more realistic to talk about professional competence, which includes linguistic competence, cultural competence, translation competence, translator competence, and technological competence. Professional competence is important for translators as it affects their career development. A remark made by Timothy Hunt is worth noting: ‘Computers will never replace translators, but translators who use computers will replace translators who don’t’ (Sofer 2009: 88). What has happened in the field of translation technology shows that Hunt’s remark may not be far off the mark. In the 1980s, very few people had any ideas about translation technology or computer-aided translation. Now, SDL alone has more than 180,000 computer-aided translators. The total number of computer-aided translators in the world is likely to be several times higher than the SDL translators.

Profit-seeking to increase productivity

Translation is in part vocational, in part academic. In the training of translators, there are courses on translation skills to foster their professionalism, and there are courses on translation theories to enhance their academic knowledge. But there are very few courses on translation as a business or as an industry. It should be noted that translation in recent decades has increasingly become a field of entrepreneurial humanities as a result of the creation of the project management function in computer-aided translation systems. This means translation is now a field of humanities which is entrepreneurial in nature. Translation as a commercial activity has to increase productivity to make more profits.

Labour-saving to increase productivity

Computer-aided translation systems help to increase productivity and profits through labour-saving, eliminating repetitive translation tasks. Through reusing past translations, an enormous amount of labour is saved. Computer-aided translation tools support translators by freeing them from boring work and letting them concentrate on what they can do best over machines, i.e. handling semantics and pragmatics. Generally, this leads to a broader acceptance by translators. The role of a translator, therefore, has changed drastically in the modern age of digital communication. Rather than simply translating the document, a computer-aided translator has to engage in other types of work, such as authoring, pre-editing, interactive editing, post-editing, term database management, translation memory database management, text alignment and manual alignment verification. It is estimated that with the use of translation technology, the work that was originally borne by six translators can be taken up by just one.
Cost-saving to increase productivity

Computer-aided translation is also cost-saving. It helps to keep the overhead cost down as what has been translated needs not to be translated again. It helps to improve budget planning.

Other issues relating to cost should also be taken into account. First, the actual cost of the tool and its periodic upgrades. Second, the licensing policy of the system, which is about the ease of transferring licences between computers or servers, the incurring of extra charges for client licences, the lending of licences to one’s vendors, freelances, and the eligibility for free upgrades. Third, the cost that is required for support, maintenance, or training. Fourth, the affordability of the system for one’s translators. Fifth, the user-friendliness of the system to one’s computer technicians and translators, which affects the cost of production.

Compatibility

The concept of compatibility in translation technology must be considered in terms of file formats, operating systems, intersystem formats, translation memory databases, terminology databases, and the languages supported by different systems.

Compatibility of file formats

One of the most important concepts in translation technology is the type of data that needs to be processed, which is indicated by its format, being shown by one or several letters at the end of a filename. Filename extensions usually follow a period (dot) and indicate the type of information stored in the file. A look at some of the common file types and their file extensions shows that in translation technology, text translation is but one type of data processing, though it is the most popular one.

There are two major types of formats: general documentation types and software development types.

(I) General documentation types

(1) Text files
All computer-aided translation systems which use Microsoft Word as text editor can process all formats recognized by Microsoft Word. Throughout the development of translation technology, most computer-aided translation systems process text files (.txt). For Microsoft Word 2000−2003, text files were saved and stored as .doc (document text file/word processing file); for Microsoft Word 2007−2011, documents were saved and stored as .docx (Document text file (Microsoft Office 2007)), .dotx (Microsoft Word 2007 Document Template). Other types of text files include .txt (Text files), .txml (WordFast files), and .rtf (Rich Text files).

All automatic and interactive translation systems can process text files, provided the text processing system has been installed in the computer before processing begins. Some of the computer-aided translation systems which can only translate text files include: Across, AidTransStudio, Anaphraseus (formerly known as OpenWordfast), AnyMem (.docx or higher), Araya, Autshumato Integrated Translation Environment (ITE), CafeTran, Déjà Vu, Esperantilo, Fluency, Fusion, OmegaT, Wordfast, and WordFisher. Computer-aided translation systems which can translate text files as well as other formats include CafeTran, Esperantilo, Felix, Fortis, GlobalSight, Google Translator Toolkit, Heartsome Translation Suite, Huajian IAT, Lingo, Lingotek, MadCap Lingo, MemoQ, MemOrg, MemSource, MetaTexis, MultiTrans,
Computer-aided translation

OmegaT+, Pootle, SDL-Trados, Similis, Snowman, Swordfish, TM-database, Transit, Wordfast, XTM, and Yaxin.

(2) Web-page files
HyperText Markup Language (HTML) is a markup language that web browsers use to interpret and compose text, images and other material into visual or audible web pages. HTML defines the structure and layout of a web page or document by using a variety of tags and attributes. HTML documents are stored as .asp (Active Server Pages), .aspx (Active Server Page Extended), .htm (Hypertext Markup Language), .html (Hypertext Markup Language Files), .php (originally: Personal Home Page; now: Hypertext Preprocessor), .jsp (JavaServer Pages), .sgml (Standard Generalized Markup Language File), .xml (Extensible Markup Language file), .xsl (Extensible Stylesheet Language file) files format, which were available since late 1991. Due to the popularity of web pages, web translation has been an important part of automatic and interactive translation systems. Many systems provide comprehensive support for the localization of HTML-based document types. Web page localization is interchangeable with web translation or web localization.

Systems that handle HTML include Across, AidTransStudio, Alchemy Publisher, Araya, Atlas, CafeTran, CatsCradle, Déjà Vu, Felix, Fluency, Fortis, GlobalSight, Google Translator Toolkit, Heartsome Translation Suite, Huajian IAT, Lingo, Lingotek, LogiTerm, MemoQ, MemOrg, MetaTexis, MultiTrans, Okapi Framework, OmegaT, OmegaT+, Open Language Tools, Pootle, SDL-Trados, Similis, Snowman, Swordfish, TM-database, TransSearch, Transit, Transolution, and XTM.

(3) PDF files
Portable Document Format (PDF) (.pdf) is a universally accepted file interchange format developed by Adobe in the 1990s. The software that allows document files to be transferred between different types of computers is Adobe Acrobat. A PDF file can be opened by the document format, which might require editing to make the file look more like the original, or can be converted to an .rtf file for data processing by a computer-aided translation system.

Systems that can translate Adobe PDF files and save them as Microsoft Word documents include Alchemy Publisher, CafeTran, Lingo, Similis, and Snowman.

(4) Microsoft Office PowerPoint files
Microsoft PowerPoint is a presentation program developed to enable users to create anything from basic slide shows to complex presentations, which are comprised of slides that may contain text, images, and other media. Versions of Microsoft Office PowerPoint include Microsoft PowerPoint 2000–2003, .ppt (General file extension), .pps (PowerPoint Slideshow), .pot (PowerPoint template); Microsoft PowerPoint 2007/2011, which are saved as .pptx (Microsoft PowerPoint Open XML Document), .ppsx (PowerPoint Open XML Slide Show), .potx (PowerPoint Open XML Presentation Template), and .ppsm (PowerPoint 2007 Macro-enabled Slide Show).

Systems that can handle PowerPoint files include Across, AidTransStudio, Alchemy Publisher, CafeTran, Déjà Vu, Felix, Fluency, Fusion, GlobalSight, Lingotek, LogiTerm, MadCap Lingo, MemoQ, MemSource, MetaTexis, SDL-Trados, Swordfish, TM-database, Transit, Wordfast, XTM, and Yaxin.
(5) Microsoft Excel files

The computer-aided translation systems that can translate Excel files include Across, AidTransStudio, Dèjà Vu, Felix, GlobalSight, Lingotek, LogiTerm, and MemoQ, MemOrg, MetaTexis, MultiTrans, Snowman, Wordfast, and Yaxin.

(6) Microsoft Access files
One of the computer-aided translation systems which can handle Access with .accdb (Access 2007–2010) file extension is Dèjà Vu.

(7) Image files
The processing of image data, mainly graphics and pictures, is important in computer-aided translation. The data is stored as .bmp (bitmap image file), .jpg (Joint Photographic Experts Group), and .gif (Graphics Interchange Format). One of the computer-aided translation systems that is capable of handling images is CafeTran.

(8) Subtitle files
One of the most popular subtitle files on the market is .srt (SubRip Text). OmegaT is one of the computer-aided systems that supports subtitle files.

(9) Adobe InDesign files
Adobe InDesign is desktop publishing software. It can be translated without the need of any third party software by Alchemy Publisher and AnyMem. For Alchemy Publisher, the .indd file must be exported to an .inx format before it can be processed. Other computer-aided translation systems that support Adobe InDesign files include Across, Dèjà Vu, Fortis, GlobalSight, Heartsome Translation Suite, Okapi Framework, MemoQ, MultiTrans, SDL-Trados, Swordfish, Transit, and XTM.

(10) Adobe FrameMaker Files
Adobe FrameMaker is an authoring and publishing solution for XML. FrameMaker files, .fm, .mif and .book, can be opened directly by a system if it is installed with Adobe FrameMaker.

Computer-aided translation systems that can translate Adobe FrameMaker files include Across, Alchemy Publisher (which requires a PPF created by Adobe FrameMaker before translating it. Alchemy Publisher supports FrameMaker 5.0, 6.0, 7.0, 8.0, 9.0, FrameBuilder 4.0, and FrameMaker + sgml), CafeTran, Dèjà Vu, Fortis, GlobalSight, Heartsome Translation Suite, Lingo, Lingotek, MadCap Lingo, MemoQ, MetaTexis, MultiTrans, SDL-Trados, Swordfish, Transit, Wordfast, and XTM.

(11) Adobe PageMaker files
Systems that support Adobe PageMaker 6.5 and 7 files include Dèjà Vu, GlobalSight, MetaTexis, and Transit.

(12) AutoCAD files
AutoCAD, developed and first released by Autodesk, Inc. in December 1982, is a software application for computer-aided design (CAD) and drafting which supports both 2D and 3D
formats. This software is now used internationally as the most popular drafting tool for a range of industries, most commonly in architecture and engineering.

Computer-aided translation systems that support AutoCad are CafeTran, Transit, and TranslateCAD.

(13) **DTP tagged text files**
DTP stands for Desktop Publishing. A popular desktop publishing system is QuarkXPress.

Systems that support desktop publishing include Across, CafeTran, Déjà Vu, Fortis, GlobalSight, MetaTexis, MultiTrans, SDL-Trados, and Transit.

(14) **Localization files**
Localization files include files with the standardized format for localization .xliff (XML Localization Interchange File Format) files, .ttx (XML font file format) files, and .po (Portable Object).

Computer-aided translation systems which process XLIFF files include Across Language Server, Araya, CafeTran, Esperantilo, Fluency, Fortis, GTranslator, Heartsome Translation Suite, MadCap Lingo, Lingotek, MemoQ, Okapi Framework, Open Language Tools, Poedit, Pootle, Swordfish, Transolution, Virtaal, and XTM.

(II) **Software development types**

(1) **Java Properties files**
Java Properties files are simple text files that are used in Java applications. The file extension of Java Properties file is .properties.

Computer-aided translation systems that support Java Properties File include Déjà Vu, Fortis, Heartsome Translation Suite, Lingotek, Okapi Framework, OmegaT+, Open Language Tools, Pootle, Swordfish, and XTM.

(2) **OpenOffice.org/StarOffice**
StarOffice of the Star Division was a German company that ran from 1984 to 1999. It was succeeded by OpenOffice.org, an open-sourced version of StarOffice owned by Sun Microsystems (1999–2009) and by Oracle Corporation (2010–2011), which ran from 1999–2011. Currently it is Apache OpenOffice. The format of OpenOffice is .odf (Open Document Format).

Computer-aided translation systems which process this type of file include AidTransStudio, Anaphraseus, CafeTran, Déjà Vu, Heartsome Translation Suite, Lingotek, OmegaT, OmegaT+, Open Language Tools, Pootle, Similis, Swordfish, Transolution, and XTM.

(3) **Windows resource files**
These are simple script files containing startup instructions for an application program, usually a text file containing commands that are compiled into binary files such as .exe and .dll. File extensions include .rc (Record Columnar File), .resx (NET XML Resource Template). Computer-aided translation systems that process this type of files include Across, Déjà Vu, Fortis, Lingotek, MetaTexis, and Okapi Framework.
Compatibility of operating systems

One of the most important factors which determined the course of development of computer-aided translation systems is their compatibility with the current operating systems on the market. It is therefore essential to examine the major operating systems running from the beginning of computer-aided translation in 1988 to the present, which include, among others, the Windows of Microsoft and the OS of Macintosh.

Microsoft Operating Systems

In the world of computing, Microsoft Windows has been the dominant operating system. From the 1981 to the 1995, the x86-based MS-DOS (Microsoft Disk Operating System) was the most commonly used system, especially for IBM PC compatible personal computers. Trados’s Translator’s Workbench II, developed in 1992, is a typical example of a computer-aided translation system working on DOS.

DOS was supplemented by Microsoft Windows 1.0, a 16-bit graphical operating environment, released on 20 November 1985 (Windows 2012). In November 1987, Windows 1.0 was succeeded by Windows 2.0, which existed till 2001. Déjà Vu 1.0, released in 1993, was one of the systems compatible with Windows 2.0. Windows 2.0 was supplemented by Windows 286 and Windows 386.

Then came Windows 3.0, succeeding Windows 2.1x. Windows 3.0, with a graphical environment, is the third major release of Microsoft Windows, and was released on 22 May 1990. With a significantly revamped user interface and technical improvements, Windows 3 became the first widely successful version of Windows and a rival to Apple Macintosh and the Commodore Amiga on the GUI front. It was followed by Windows 3.1x. During its lifespan from 1992–2001, Windows 3.1x introduced various enhancements to the still MS-DOS-based platform, including improved system stability, expanded support for multimedia, TrueType fonts, and workgroup networking. Trados’s Translator’s Workbench, released in 1994, was a system that was adaptable to Windows 3.1x.

Except for Windows and DOS, OS/2 is also one of the operation systems that support computer-aided translation systems, especially in late 1980s and early 1990s.

Apple Operating Systems

Mac OS (1984–2000) and OS X (2001–) are two series of graphical user interface-based operating systems developed by Apple Inc. for their Macintosh line of computer systems. Mac OS was first introduced in 1984 with the original Macintosh and this series was ended in 2000. OS X, first released in March 2001, is a series of Unix-based graphical interface operating systems. Both series share a general interface design, but have very different internal architectures.

Only one computer-aided translation system, AppleTrans, is designed for OS X. Its initial released was announced in February 2004 and the latest updated version was version 1.2(v38) released in September 2006, which runs on Mac OS X 10.3 or later.

Another computer-aided translation system, Wordfast Classic was released to upgrade its support of the latest text processor running on Mac OS X, such as Wordfast Classic 6.0, which is compatible for MS Word 2011 for Mac.

Other computer-aided translation systems that can run on Mac OS or OS X are cross-platform software, rather than software developed particularly for Mac. Examples are Java-based applications, such as Autshumato, Heartsome, OmegaT, Open Language Tools and
Swordfish. Besides, all cloud-based systems can support Mac OS and OS X, including Wordbee, XTM Cloud, Google Translator’s Toolkit, Lingotek Collaborative Translation Platform, MemSource Cloud, and WebWordSystem.

OS/2 is a series of computer operating systems, initially created by Microsoft and IBM, then later developed by IBM exclusively. The name stands for ‘Operating System/2’.

Until 1992, the early computer-aided translation systems ran either on MS-DOS or OS/2. For example, IBM Translation Manager/2 (TM/2) was released in 1992 and run on OS/2. ALPS’s translation tool also ran on OS/2. But OS/2 had a much smaller market share compared with Windows in early 1990s. Computer-aided translation system developers therefore gradually shifted from OS/2 and MS-DOS to Windows or discontinued the development of OS/2 and MS-DOS compatible computer-aided translation systems. By the end of the 1990s, most computer-aided translation systems mainly ran on Windows, although some developers offered operating-system customization services upon request. OS/2 4.52 was released in December 2001. IBM ended its support to OS/2 on 31 December 2006.

**Compatibility of databases**

**Compatibility of translation memory databases**

TMX (Translation Memory eXchange), created in 1998, is widely used as an interchange format between different translation memory formats. TMX files are XML (eXtensible Markup Language) files whose format was originally developed and maintained by OSCAR (Open Standards for Container/Content Allowing Re-use) of the Localization Industry Standards Association. The latest official version of the TMX specification, version 1.4b, was released in 2005. In March 2011 LISA was declared insolvent; as a result its standards were moved under the Creative Commons licence and the standards specification relocated. The technical specification and a sample document of TMX can be found on the website of The Globalization and Localization Association.

TMX has been widely adopted and is supported by more than half of the current computer-aided translation systems on the market. The total number of computer-aided translation systems that can import and export translation memories in TMX format is 54, including Across, Alchemy Publisher, Anaphraseus, AnyMem, Araya, ATLAS, Autshumato, CafeTran, Crowdin, Déjà Vu, EsperantiloTM, Felix, Fluency, Fortis, Fusion, GE-CCT, GlobalSight, Google Translator Toolkit, Heartsome, Huajian IAT, Lingotek, LogiTm, LongRay CAT, MadCap Lingo, MemoQ, MemSource, MetaTexis, MT2007, MultiTrans, OmegaT, OmegaT+, Open Language Tools, OpenTM2, OpenTMS, PROMT, SDL Trados, Snowball, Snowman, Swordfish, Systran, Text United, The Hongyahu, TM Database, Transit, Translation Workspace, Transwhiz, TraTool, Webwordsystem, Wordbee Translator, Wordfast Classic and Wordfast Pro, XTM, Yaxin CAT, and 翻訳ブレイン (Translation Brain).

**Compatibility of terminology databases**

Compatibility of terminology databases is best illustrated by TermBase eXchange (TBX), which covers a family of formats for representing the information in a high-end termbase in a neutral intermediate format in a manner compliant with the Terminological Markup Framework (TMF) (Melby 2012: 19–21).

Termbase Exchange is an international standard as well as an industry standard. The industry standard version differs from the ISO standard only by having different title pages. Localization
Industry Standards Association, the host organization for OSCAR that developed Termbase Exchange, was dissolved in February 2011. In September 2011, the European Telecommunications Standards Institute (ETSI) took over maintenance of the OSCAR standards. ETSI has established an interest group for translation/localization standards and a liaison relationship with the International Organization for Standardization (ISO) so that TBX can continue to be published as both an ISO standard and an industry standard.

There are many types of termbases in use, ranging from huge termbases operated by governments, to medium-size termbases maintained by corporations and non-governmental organizations, to smaller termbases maintained by translation service providers and individual translators. The problem addressed by the designers of term exchange was that existing termbases are generally not interoperable. They are based on different data models that use a variety of data categories. And even if the same data category is used for a particular piece of information, the name of the data category and the values allowed for the data category may be different.

Compatiblity of rules

SEGMENTATION RULES EXCHANGE

Segmentation Rules eXchange (SRX) is an XML-based standard that was maintained by Localization Industry Standards Association, until it became insolvent in 2011 and then this standard is now maintained by the Globalization and Localization Association (GALA).

Segmentation Rules eXchange provides a common way to describe how to segment text for translation and other language-related processes. It was created when it was realized that translation memory exchange leverage is lower than expected in certain instances due to differences in how tools segment text. Segmentation Rules eXchange is intended to enhance the translation memory exchange so that translation memory data that is exchanged between applications can be used more effectively. Having the segmentation rules that were used when a translation memory was created will increase the leverage that can be achieved when deploying the translation memory data.

Compatibility with the languages supported

As computer-aided translation systems cannot identify languages, language compatibility is therefore an important concept in translation technology. There are a large number of languages and sub-languages in the world, totalling 6,912. But the number of major languages computers can process is relatively small. It is therefore important to know whether the languages that require machine processing are supported by a system or not.

With the aid of unicode, most of the languages in the world are supported in popular computer-aided translation systems. Unicode is a computing industry standard for the consistent encoding, representation and handling of text expressed in most of the world’s writing systems.

There are basically two major types of language and sub-language codes. Some systems, such as OmegaT and XTM, use letters for language codes (2 or 3 letters) and language-and-region codes (2+2 letters), which can be selected from a drop-down list. OmegaT follows the ISO 639 Code Tables in preparing its code list. French for example, is coded fr with the language-and-region code for French (Canada) as fr-CA.

The following is a list of languages supported by Wordfast Classics and XTM, two of the nine computer-aided translation systems chosen for analysis in this chapter.
Computer-aided translation

Wordfast can be used to translate any of the languages supported by Microsoft Word. The number of languages supported by Microsoft is 91, with a number of sub-languages for some major languages.

*Afro-Asiatic* Arabic (Algeria), Arabic (Bahrain), Arabic (Egypt), Arabic (Iraq), Arabic (Jordan), Arabic (Kuwait), Arabic (Lebanon), Arabic (Libya), Arabic (Morocco), Arabic (Oman), Arabic (Qatar), Arabic (Saudi Arabia), Arabic (Syria), Arabic (Tunisia), Arabic (U.A.E.), Arabic (Yemen), Hebrew, Maltese

*Altaic* Azeri (Cyrillic), Azeri (Latin), Japanese, Korean, Turkish

*Austro-Asiatic* Vietnamese

*Austronesian* Indonesian, Malay (Brunei Darussalam), Malaysian

*Basque* Basque

*Dravidian* Kannada, Malayalam, Tamil, Telugu

*Indo-European* Afrikaans, Albanian, Armenian, Assamese, Belarusian, Bengali, Bulgarian, Byelorussian, Catalan, Croatian, Czech, Danish, Dutch, Dutch (Belgian), English (Australia), English (Belize), English (Canadian), English (Caribbean), English (Ireland), English (Jamaica), English (New Zealand), English (Philippines), English (South Africa), English (Trinidad), English (U.K.), English (U.S.), English (Zimbabwe), Faroese, Farsi, French (Belgian), French (Cameroon), French (Canadian), French (Congo), French (Cote d’Ivoire), French (Luxembourg), French (Mali), French (Monaco), French (Reunion), French (Senegal), French (West Indies), Friisian (Netherlands), Gaelic (Ireland), Gaelic (Scotland), Galician, German, German (Austria), German (Liechtenstein), German (Luxembourg), Greek, Gujarati, Hindi, Icelandic, Italian, Kashmiri, Konkani, Latvian, Lithuanian, Macedonian (FYRO), Marathi, Nepali, Norwegian (Bokmål), Norwegian (Nynorsk), Oriya, Polish, Portuguese, Portuguese (Brazil), Punjabi, Rhaeto-Romance, Romanian, Romanian (Moldova), Russian, Russian (Moldova), Sanskrit, Serbian (Cyrillic), Serbian (Latin), Sindhi, Slovak, Slovenian, Sorbian, Spanish (Argentina), Spanish (Bolivia), Spanish (Chile), Spanish (Colombia), Spanish (Costa Rica), Spanish (Dominican Republic), Spanish (Ecuador), Spanish (El Salvador), Spanish (Guatemala), Spanish (Honduras), Spanish (Nicaragua), Spanish (Panama), Spanish (Paraguay), Spanish (Peru), Spanish (Puerto Rico), Spanish (Spain), Spanish (Traditional), Spanish (Uruguay), Spanish (Venezuela), Swedish, Swedish (Finland), Swiss (French), Swiss (German), Swiss (Italian), Tajik, Ukrainian, Urdu, Welsh

*Kartvelian* Georgian

*Niger-Congo* Sesotho, Swahili, Tsonga, Tswana, Venda, Xhosa, Zulu

*Sino-Tibetan* Burmese, Chinese, Chinese (Hong Kong SAR), Chinese (Macau SAR), Chinese (Simplified), Chinese (Singapore), Chinese (Traditional), Manipuri, Tibetan
The languages available in XTM are 157, not including varieties within a single language. These languages are as follows:

**Afro-Asiatic** Afar, Amharic, Arabic, Hausa, Hebrew, Maltese, Oromo, Somali, Sudanese Arabic, Syriac, Tigrinya,

**Altaic** Azeri, Japanese, Kazakh, Korean, Mongolian, Turkish

**Austro-Asiatic** Khmer, Vietnamese

**Austronesian** Fijian, Indonesian, Javanese, Malagasy, Malay, Maori, Nauru, Samoan, Tagalog, Tetum, Tonga

**Aymaran** Aymara

**Bantu** Kikongo

**Basque** Basque

**Constructed Language** Esperanto, Interlingua, Volapk

**Dravidian** Kannada, Malayalam, Tamil, Telugu

**English Creole** Bislama

**Eskimo-Aleut** Greenlandic, Inuktitut, Inupiak

**French Creole** Haitian Creole

**Hmong-Mien** Hmong

**Indo-European** Afrikaans, Armenian, Assamese, Asturian, Bengali, Bihari, Bosnian, Breton, Bulgarian, Byelorussian, Catalan, Corsican, Croatian, Czech, Danish, Dari, Dhivehi, Dutch, English, Faroese, Flemish, French, Frisian, Galician, German, Greek, Gujarati, Hindi, Icelandic, Irish, Italian, Kashmiri, Konkani, Kurdish, Latin, Latvian, Lithuanian, Macedonian, Marathi, Montenegrin, Nepali, Norwegian, Occitan, Oriya, Pashto, Persian, Polish, Portuguese, Punjabi, Rhaeto-Romance, Romanian, Russian, Sanskrit, Sardinian, Scottish Gaelic, Serbian, Sindhi, Singhalese, Slovak, Slovenian, Sorbian, Spanish, Swedish, Tajik, Ukrainian, Urdu, Welsh, Yiddish
Several observations can be made from the languages supported by the current eleven systems.

1. The number of languages supported by language-specific systems is small as they need to be supplied with language-specific dictionaries to function well. Yaxin is best for English–Chinese translation, covering two languages, while most non-language-specific systems support around or above 100 languages.

2. For the seven systems developed in Europe, the United Kingdom, and the United States, which include Across, Déjà Vu, MemoQ, OmegaT, SDL Trados, Wordfast, and XTM, the Indo-European languages take up around 51.89 per cent, while the proportion of the non-Indo-European languages is 48.11 per cent. Table 2.1 shows the details:

Table 2.1 Statistics of languages supported by 7 CAT systems

<table>
<thead>
<tr>
<th>Name of the system</th>
<th>Number of languages supported</th>
<th>Number of language families supported</th>
<th>Number and percentage of Indo-European languages</th>
<th>Number and percentage of non-Indo-European languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across</td>
<td>121</td>
<td>18</td>
<td>61 (50.41%)</td>
<td>60 (49.59%)</td>
</tr>
<tr>
<td>Déjà Vu</td>
<td>132</td>
<td>21</td>
<td>66 (50%)</td>
<td>66 (50%)</td>
</tr>
<tr>
<td>MemoQ</td>
<td>102</td>
<td>16</td>
<td>54 (52.94%)</td>
<td>48 (47.06%)</td>
</tr>
<tr>
<td>OmegaT</td>
<td>90</td>
<td>14</td>
<td>48 (53.33%)</td>
<td>42 (46.67%)</td>
</tr>
<tr>
<td>SDL Trados</td>
<td>115</td>
<td>18</td>
<td>62 (53.91%)</td>
<td>53 (46.09%)</td>
</tr>
<tr>
<td>Wordfast</td>
<td>91</td>
<td>13</td>
<td>54 (59.34%)</td>
<td>37 (40.66%)</td>
</tr>
<tr>
<td>XTM</td>
<td>157</td>
<td>26</td>
<td>68 (43.31%)</td>
<td>89 (56.69%)</td>
</tr>
</tbody>
</table>
Controllability

One of the main differences between human translation and computer-aided translation lies in the degree of control over the source text. In human translation, there is no need, or rather it is not the common practice, to control how and what the author should write. But in computer-aided translation, control over the input text may not be inappropriate as the output of an unedited or uncontrolled source language text is far from satisfactory (Adriaens and Macken 1995: 123–141; Allen and Hogan 2000: 62–71; Arnold et al. 1994; Hurst 1997: 59–70; Lehtola, Tenni and Bounsaythip 1998: 16–29; Mitamura 1999: 46–52; Murphy et al. 1998; Nyberg et al. 2003: 245–281; Ruffino 1985: 157–162).

The concept of controllability is realized in computer-aided translation by the use of controlled language and the method of pre-editing.

Controllability by the use of controlled language

An effective means of achieving controllability in translation technology is controlled language (see Figure 2.16). The idea of controlled language was created, partly at least, as a result of the problems with natural languages which are full of complexities, ambiguities, and robustness (Nyberg et al. 2003: 245–281). A strong rationale for controlled language is that a varied source text generates a poor target text, while a controlled source text produces a quality target text. (Bernth 1999). Controlled language is therefore considered necessary (Caeyers 1997: 91–103; Hu 2005: 364–372).

Controlled language, in brief, refers to a type of natural language developed for specific domains with a clearly defined restriction on controlled lexicons, simplified grammars, and style rules to reduce the ambiguity and complexity of a text so as to make it easier to be understood by users and non-native speakers and processed by machine translation systems (Chan 2004: 44; Lux and Dauphin 1996: 193–204).

Control over the three stages of a translation procedure, which include the stage of inputting a source text, the stage of transfer, and the stage of text generation, is generally regarded as a safe guarantee of quality translation. Control of the source text is in the form of controlled authoring, which makes the source text easier for computer processing (Allen 1999; Chan 2004: 44; van der Eijk and van Wees 1998: 65–70; Zydron 2003). The text produced is a ‘controlled language text’ (Melby 1995: 1). There is also control over the transfer stage. And the output of a machine translation system is known as ‘controlled translation’ (Carl 2003: 16–24; Gough and Way 2004: 73–81; Rico and Torrejon 2004; Roturier 2004; Torrejón 2002: 107–116), which is alternatively known as a ‘controlled target language text’ (Chan 2004: 44). In short, a controlled text is easier to be processed by machine translation systems to produce a quality output.

Goals and means of controlled language

Controlled language is used by both humans and computers. The goals of controlled language are to make the source text easier to read and understand. These goals are to be achieved at the lexical and sentential levels.

At the lexical level, controlled language is about the removal of lexical ambiguity and the reduction in homonymy, synonymy, and complexity. This is to be achieved by one-to-one correspondence in the use and translation of words, known as one-word one-meaning. An example is to use only the word ‘start’ but not similar words such as ‘begin’, ‘commence’,
Figure 2.16 Controlled language

‘initiate’, and ‘originate’. The second method is to use the preferred language, such as American English but not British English. The third method is to have a limited basic vocabulary (Bjarnestam 2003; Chen and Wu 1999; Probst and Levin 2002: 157–167; Wasson 2000: 276–281), which can be illustrated by the use of a controlled vocabulary of 3,100 words in aircraft-maintenance documentation at the European Association of Aerospace Industries (AECMA) in 1980 (AECMA 1995).

At the sentential level, controlled language is about the removal of syntactical ambiguity, the simplification of sentence structures, limitations on sentence length, and constraints on voice, tense, and other grammatical units. To do all these, there are a limited number of strictly stipulated writing rules to follow. The European Association of Aerospace Industries had 57 writing rules. Short sentences are preferred over long and complex sentences. And there is also a limit on the number of words in a sentence. For procedural text, there should be no more than twenty words. For descriptive texts, the number is twenty-five. There are also rules governing grammatical well-formedness (Loong 1989: 281–297), restricted syntax, and the use of passive construction in procedural texts. At the suprasentential level, there is a limit of six sentences in a paragraph, the maximum number of clauses in a sentence, and the use of separate sentences for sequential steps in procedural texts.

This means setting limits on the length of a sentence, such as setting the number of words at twenty, using only the active voice, and expressing one instruction or idea by one sentence.

**Controlled language checkers**

Controlled language cannot be maintained manually; it relies on the use of different kinds of checkers, which are systems to ensure that a text conforms to the rules of a particular controlled language (Fouvry and Balkan 1996: 179–192). There is the automatic rewriting system, which is specially developed for controlled language, rewriting texts automatically into controlled language without changing the meaning of the sentences in the original in order to produce a high-quality machine translation. There is the controlled language checker, which is software that helps an author to determine whether a text conforms to the approved words and writing rules of a particular controlled language.

Checkers can also be divided into two types: in-house controlled language checker and commercial controlled language checker. In-house controlled language checkers include the
PACE (Perkins Approved Clear English) of Perkins Engines Ltd, the Controlled English of Alcatel Telecom, and the Boeing Simplified English Checker of the Boeing Company (Wojcik and Holmback 1996: 22–31). For commercial controlled language checkers, there are a number of popular systems. The LANTmaster Controlled Checker, for example, is a controlled language checker developed by LANT in Belgium. It is based on work done for the METAL (Mechanical Translation and Analysis of Languages) machine translation project. It is also based on the experience of the Simplified English Grammar and Style Checker (SECC) project (Adriaens 1994: 78–88; Adriaens and Macken 1995: 123–141). The MAXit Checker is another controlled language software developed by Smart Communications Incorporation to analyse technical texts written in controlled or simplified English with the use of more than 8,500 grammar rules and artificial intelligence to check the clarity, consistency, simplicity, and global acceptance of the texts. The Carnegie Group also produced the ClearCheck, which performs syntactic parsing to detect such grammatical problems as ambiguity (Andersen 1994: 227).

Advantages and disadvantages of controlled language

The advantages of controlled language translation are numerous, including high readability, better comprehensibility, greater standardization, easier computer processing, greater reusability, increased translatability, improved consistency, improved customer satisfaction, improved competitiveness, greater cost reduction in global product support, and enhanced communication in global management.

There are a number of disadvantages in using controlled language, such as expensive system construction, high maintenance cost, time-consuming authoring, and restrictive checking process.

Controlled language in use


Controlled language in computer-aided translation systems

The concept of controlled language is realized in controlled authoring in computer-aided translation systems. Authoring checking tools are used to check and improve the quality of the source text. There is an automatic rewriting system which is usually used as a tool to realize controlled authoring. One of the computer-aided translation systems that performs controlled
authoring is Star Transit. This system provides automatic translation suggestions from the translation memory database from a speedy search engine and it is an open system that can integrate with many authoring systems.

**Customizability**

Customizability, etymologically speaking, is the ability to be customized. More specifically, it refers to the ability of a computer or computer-aided translation system to adapt itself to the needs of the user. Customizing a general-purpose machine translation system is an effective way to improve MT quality.

**Editorial customization**

Pre-editing is in essence a process of customization. The customization of machine translation systems, which is a much neglected area, is necessary and essential as most software on the market are for general uses and not for specific domains. Practically, system customization can be taken as part of the work of pre-editing as we pre-edit the words and expressions to facilitate the production of quality translation.

The degree of customization depends on the goals of translation, and the circumstances and the type of text to be translated.

**Language customization**

It is true that there are many language combinations in computer-aided translation systems to allow the user to choose any pair of source and target languages when creating a project, yet many users only work with a limited set of source and target languages. XTM, a cloud-based system, allows the user to set language combinations through the Data section. In the language combinations section, the project administrator or user can reduce and customize the available languages to be used, set the language combinations for the entire system and set specific language combinations for individual customers (XTM International 2012: 15).

Language customization in XTM, for example, can be conducted on the Customize tab where there are three options for the user to modify and use language combinations. The first option is ‘system default language combinations’, which is the full set of unmodified language combinations. The second option is ‘system defaults with customized language combinations’, which is the full set of language combinations in which the user may have customized some parameters. The third option is ‘customized language combinations only’, which include only the language combinations that the user has customized. It is possible to add or delete the source and target languages in the selected customized option.

**Lexicographical customization**

Lexicographical customization is best shown in the creation of custom dictionaries for each customer, other than the dictionaries for spell checking. This means that multiple translators working on projects for the same customer will use the same custom dictionary.
Linguistic customization

As far as linguistic customization is concerned, there are basically two levels of customization: lexical customization and syntactical customization.

Lexical customization

Lexical customization is to customize a machine translation system by preparing a customized dictionary, in addition to the system dictionary, before translating. This removes the uncertainties in translating ambiguous words or word combinations. It must be pointed out, however, that the preparation of a customized dictionary is an enormous task, involving a lot of work in database creation, database maintenance, and database management.

Syntactical customization

Syntactical customization, on the other hand, is to add sentences or phrases to the database to translate texts with many repetitions. Syntactical customization is particularly important when there is a change of location for translation consumption. The translation memory databases built up in Hong Kong for the translation of local materials, for example, may not be suitable for the production of translations targeted at non-Hong Kong readers, such as those in mainland China.

Resource customization

Website customization

Some computer-aided translation systems allow the user to create resource profile settings. Each profile in Fluency, for example, has four customized uniform resource locators (URLs) associated with it. URLs are the Internet addresses of information. Each document or file on the Internet has a unique address for its location. Fluency allows the user to have four URLs of one’s preference, two perhaps for specialized sites and two general sites.

Machine translation system customization

Some systems are connected to installed machine translation systems the terminology databases of which can be customized for the generation of output, thus achieving terminological consistency in the target text.

Collaborativity

Collaborativity is about continuously working and communicating with all parties relating to a translation project, from the client to the reviewer, in a shared work environment to generate the best benefits of team work. Computer-aided translation is a modern mode of translation production that works best in team translation. In the past and decreasingly at present, individual translation has been the norm of practice. At present and increasingly in the future, team translation is the standard practice.

A number of systems, such as Across and Wordfast, can allow users to interact with each other through the translation memory server and share translation memory assets in real time.
Translation is about management. Translation business operates on projects. Translation technology is about project management, about how work is to be completed by translation teams. With the use of translation technology, the progress of translation work is under control and completed with higher efficiency. The best way to illustrate this point is project collaboration, which allows translators and project managers to easily access and distribute projects and easily monitor their progress.

The work of translation in the present digital era is done almost entirely online with the help of a machine translation or computer-aided translation system. This can be illustrated with SDL-Trados 2014, which is a computer-aided translation system developed by SDL International and generally considered to be the most popular translation memory system on the market.

Figure 2.17 shows the dashboard of SDL-Trados 2014.

![Dashboard of SDL-Trados 2014](image)

Figure 2.17  Dashboard of SDL-Trados 2014

![List of current projects](image)

Figure 2.18  List of current projects
Workflow of a translation project

To start a project, the first stage of the workflow is the creation of a termbase and a translation memory database, as shown in Figure 2.20.

Figure 2.20  Workflow of a translation project: the first stage
In other words, when the Project Manager has any publications, files or web pages to translate, he will send them to the translators of a department or unit, or freelancers for processing. They will create translation units and term databases from these pre-translated documents and save these databases in the SDL-Trados 2014 Server. This is the first stage of the workflow.

After the creation of translation memory and term databases, as shown in Figure 2.21, the Project Manager can then initiate a translation project and monitor its progress with the use of SDL-Trados 2014 (as indicated by (1)). He can assign and distribute source files to in-house and / or freelance translators by emails (as indicated by (2)). Translators can then do the translation by (i) reusing the translation memories and terms stored in the databases; (ii) adding new words or expressions to the translation memory and term databases (as indicated by (3)). When the translation is done, translators send their translated files back to the Project Manager on or before the due date (as indicated by (4)). When the Project Manager receives the translated files, he updates the project status, finalizes the project and marks it as ‘complete’ (as indicated by (5)).

To make sure that SDL-Trados 2014 has a smooth run, a technical support unit to maintain the SDL-Trados server may be necessary (as indicated by (6)).

Figure 2.21  Workflow of a translation project: the second stage

A translation team usually consists of the following members.

**Project manager**

A project manager is a professional in the field of project management. The responsibilities of a project manager include the following:

1. plan, execute, and close projects
   (When planning a project, the project manager works on the overall resources and budget of the project. When executing a project, the project manager can add or import customers and subcontract projects.)
2. create clear and attainable project objectives;
3. build the project requirements; and
4. manage cost, time, and scope of projects.

63
Terminologist

A terminologist is one who manages terms in the terminology database. There are two types of terminologists: (1) customer-specific terminologists who can only access the terminology of one customer; and (2) global experts who can access all the terms in the systems for all customers.

Conclusion

This chapter is possibly the first attempt to analyse the concepts that have governed the growth of functionalities in computer-aided translation systems. As computing science and related disciplines advance, more concepts will be introduced and more functions will be developed accordingly. However, it is believed that most of the concepts discussed in this chapter will last for a long time.

References


http://www.helicon.co.at/aboutus.html.
http://www2.multilizer.com/company.


66