

CHAPTER II

TERMINOLOGY MANAGEMENT TOOLS – A GENERAL OVERVIEW

1. INTRODUCTION

Having been presented with the basics of terminology management as an interdisciplinary science we should now consider terminology management from a pragmatic point of view. Therefore the following chapter of the present thesis will focus on the presentation of terminology management tools, their general features and functionalities.

First, the users will be presented with the basic principles according to which the terminology management tools operate. Since most terminology management tools have a number of common characteristics, these will be discussed right after this introduction. The differences between particular tools will be illustrated later, in the testing procedure applied to the selected programs. The features of the tools discussed shall be presented in the order in which the evaluation procedure will later be applied. Thus, first the readers will find an overview of the general software and hardware requirements of the tools including compatibility issues. Next, some details of the user interface shall be presented along with on-screen display. Later the issues of data management will be discussed, followed by a thorough discussion of entry models and structures. As a next step, the issues of data retrieval and access will be detailed. The next section shall be devoted to the systems' responses to queries, security of information, data input, terminology extraction, validation and control. Also the issues of data exchange, import and export will be presented briefly. Finally, the discussion will focus on the interaction of the terminology management tools with other applications such as word processors, translation memories and other. Also some attention will be devoted to fonts and character sets and maintenance operations. Commercial aspects of the tools under investigation will be discussed as the last step of the presentation.

Before applying the evaluation procedure, the author decided to remark upon the advantages and some possible limitations of the particular functionalities and solutions. A long-term membership of the mailing lists of users of the tools later subject to evaluation helped the author to monitor the problems and inconveniences the professional users face, as well as acquire a broader understanding of the tools' advantages. The comments resulting from this research seem to be a good complement to the above presentation, and as such will

constitute a good background for conducting the evaluation procedure in chapter IV of the present thesis.

2. TERMINOLOGY MANAGEMENT TOOLS IN THE TRANSLATION PROCESS

Terminology management tools act as specialist machine-readable dictionaries equipped with a number of additional functionalities. The termbases may be populated manually by the translators, either at the beginning of a project, or while actually performing the translation. They can also be fed with the existing terminology collections by way of import. The difference between a regular machine-readable dictionary and a TMS is that terminology management tools allow not only for the creation of custom dictionaries, with a user-definable entry structure, but also that these tools offer a number of automated functions, e.g. automatic term recognition and insertion, quality checks, etc.

The typical workflow for translators using workbench tools consists of a number of phases, each involving both TM and TMS modules. The first phase – preparation consists in preparing all the modules for a given translation project. It may require importing external translation memories and terminology collections into respective modules, and, depending on the tool, also importing the SL files into the project⁷. Depending on the resources provided by the client, it is also possible to create translation memories and automatic terminology extraction by way of alignment. Once the data is imported to the modules, translation memory and termbase should be associated with the project. Pretranslation is the next step. This function divides the source file into segments and searches through the termbase and translation memory, looking for the matching terms and segments. Once the analysis is done and the target segments have been filled with what was found in the databases, the actual translation may begin. During translation the terminology database can be searched and the terminological records can be consulted in order to make the best possible choice. New terms can be added throughout the whole process, either by way of automatic sending the term pairs to the termbase or by manual data entry. Also translation memory is updated during translation. Finally, all kinds of quality checks are performed, including spell checking, numbers and terminology checks i.e. whether all instances of a

⁷ Some applications require importing the SL file into the application. The translation process is then carried out in the native format of the tool. Other applications act as macros associated with word processors and do not require import, but operate directly on the original file.

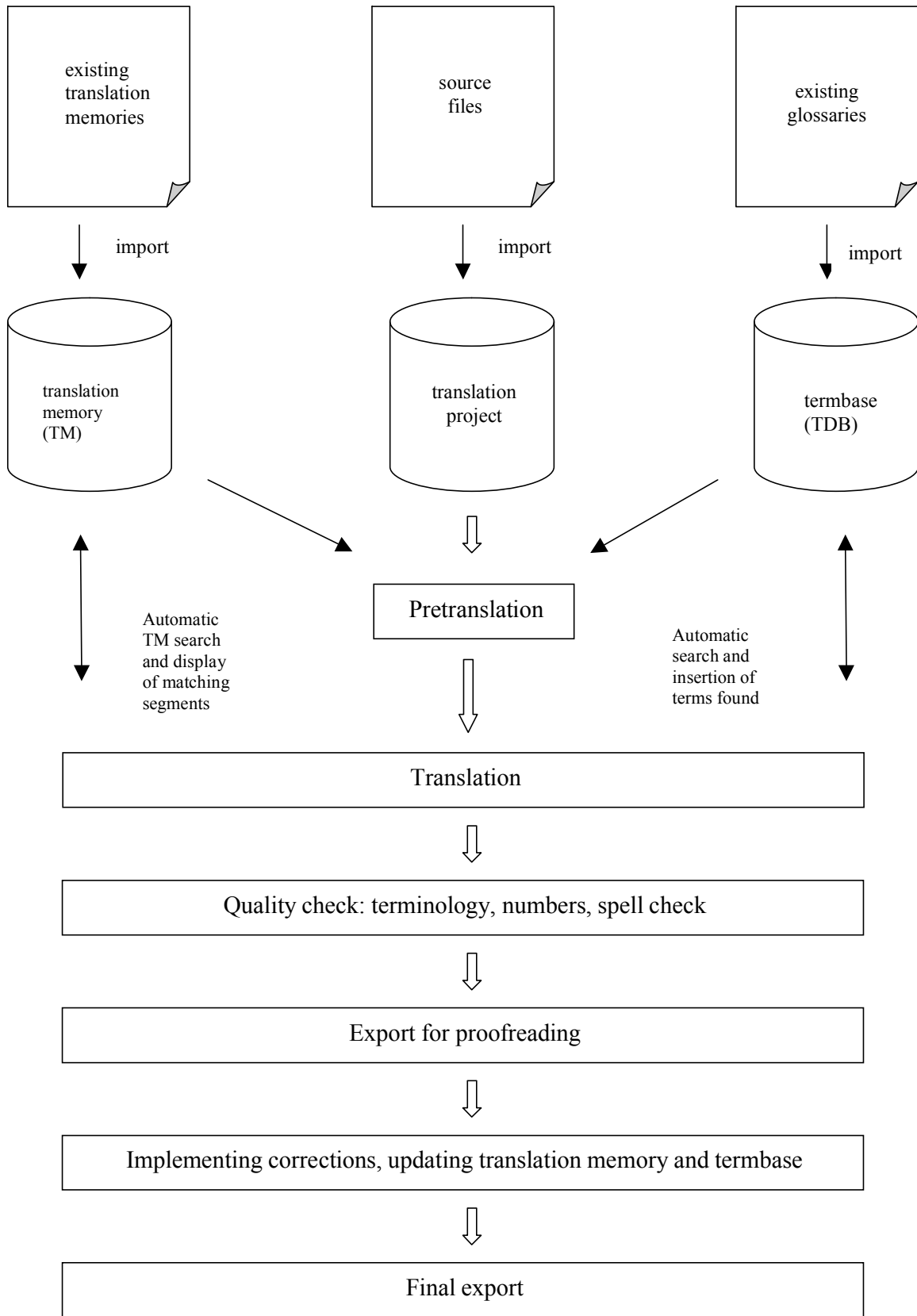


Diagram 1. Usual workflow model in translation using TM and TMS modules.

given SL term have been translated in the same way into TL and all numbers have been converted into the appropriate format. Next, the text can be exported for proofreading, following which, both termbase and translation memory should be updated. The final export is the last step of the CAT-supported translation process. The above diagram prepared by the author illustrates this process. A more specific description of the functionalities outlined in this process description will be provided in the next section.

3. TERMINOLOGY MANAGEMENT TOOLS – FUNCTIONALITIES

3.1 Software and hardware requirements

Most terminology management tools are designed for use on PC computers. However, they have specific hardware requirements that need to be met for the software to be fully functional. Therefore, bearing in mind the high prices of terminology management tools, it is necessary to check whether a computer intended for installation of this software fulfills the requirements of hard disc space, RAM size and processor type. Hardware requirements additionally include such criteria as minimum screen resolution, or graphics card. Finally, terminology management tools, like most other contemporary software require such peripheral devices as mouse or trackball, CD-ROM, printer, monitor, and in the case of multi-user versions, also a network card. If the tool comes also as a client-server application, it is necessary to have a modem.

Apart from hardware concerns, it is essential to know what additional software is required so that the programs could run successfully. Before installing any of the terminology management tools the support of the operating system installed on the computer must be confirmed. It is also worth knowing if the program is multi-user enabled. In most cases terminology management tools can be bought in free-lance versions designed for a single user and network versions for translation agencies.

A very important issue is multitasking. All modern operating systems applying graphic user interface (GUI) ensure full multitasking, however it is worth checking whether the tool itself allows for a number of processes to be run simultaneously. As Feder explains: ‘A user might find it necessary to run two types of searches at the same time or work on translating a given segment and look for occurrence of similar phrases [...]’ (Feder 2001:139).

Another question of major concern is what additional software is required for the tool to run successfully. Most terminology management applications require MS Word, MS Excel and MS PowerPoint to be installed in order to be in full working order. In some cases also MS Internet Explorer is required.

3.2 Compatibility

The term *compatibility* refers to two major issues. It denotes the compatibility of the software with the operating system and other additional applications required on the one hand, and between the different versions of the same tool on the other. Since compatibility of the tool with the operating system is probably one of the first things verified by the purchaser, in most cases it does not cause major problems. The support of the operating systems and the recommended versions are clearly indicated in the documentation. Another issue is the compatibility of the different versions of the tool. It is frequently so that the new versions do not support the format of the early versions of the same software. It is therefore an extremely important criterion in making up one's mind whether to upgrade the tool or not. However, if there is no back compatibility between the versions, the new version of a program is usually equipped with a functionality which enables converting the files created in the old format into the new one, e.g. in the case of Multiterm *iX* a module called MultitermConvert easily converts the files created in MultiTerm 5.x into Multiterm *iX* format. Déjà vu X, in turn, has a similar functionality in the menu under Tools>Convert. Finally, if an old terminology database cannot be used in a new version of the program, it is often possible to export the contents of the termbase into an intermediary format, text or spreadsheet, and import into the native format⁸ of the new version of the tool.

3.3 User interface

User interface is the means by which the user can communicate with the computer and its applications. Usually it involves such elements as menus, typed commands, function keys or mouse clicks. The first generation of interfaces had the inherent limitations e.g. hindering the simultaneous accessibility of termbase entries for editing and reference (Fontenelle&Mergen 1998). Nowadays, virtually all personal computers are equipped with the GUI consisting of windows, panes, bars, buttons, icons and menus. The same is true for the modern terminology management tools. However, there are sometimes significant

⁸ Native format – the file format characteristic to a given terminology management or translation memory tool (EAGLES 1995: 141).

differences between the applications in the way the program communicates with the user and displays data. Figure 1 represents the interface of Multiterm *iX*, while Figure 2 shows the interface of DVX.

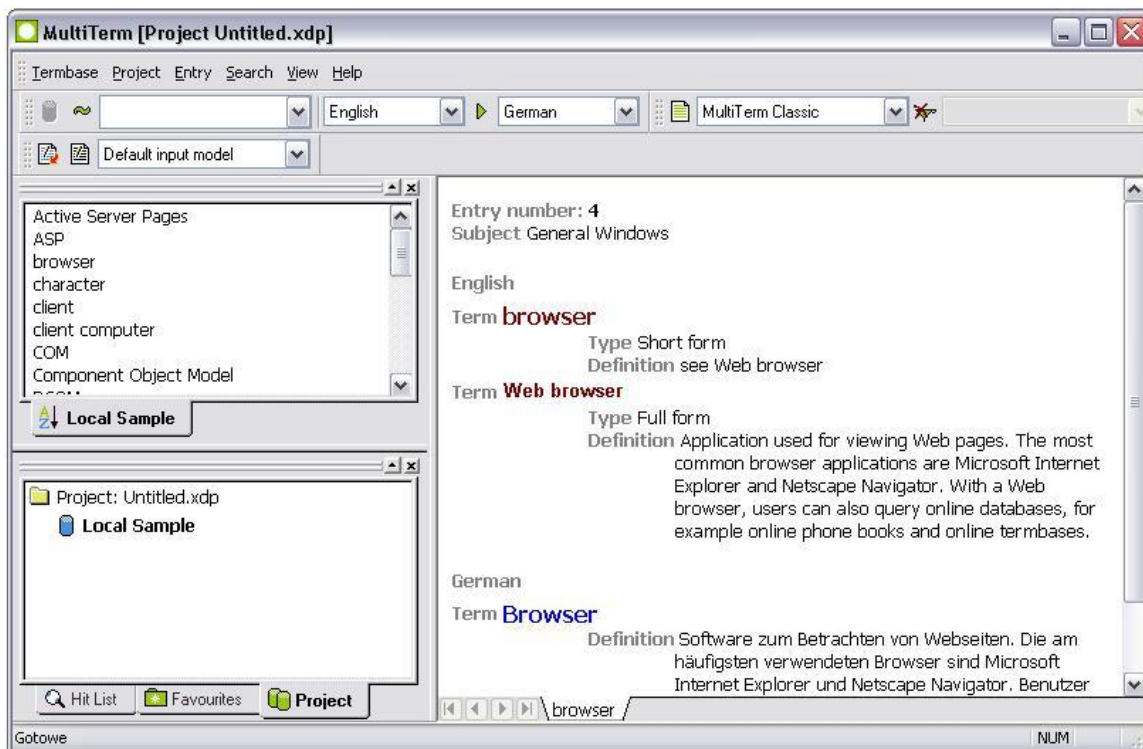


Figure 1. Multiterm *iX* interface – sample termbase provided with the application.

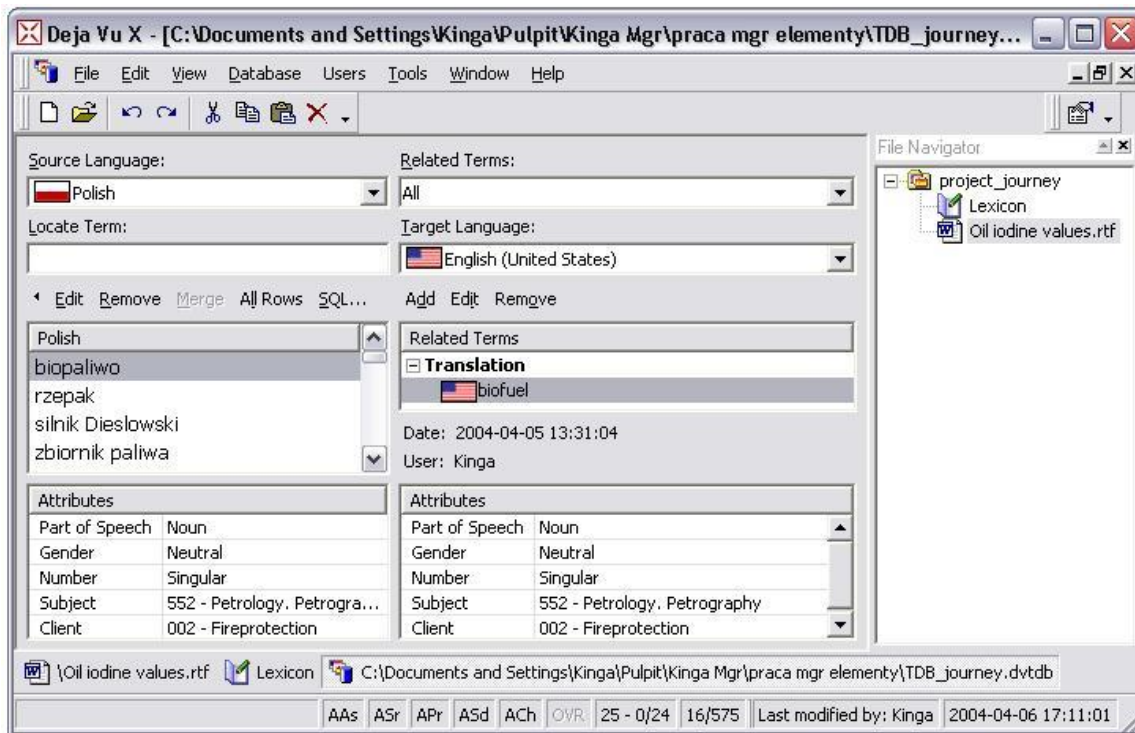


Figure 2. Déjà vu X interface – termbase created for the purpose of this project.

As we can see, even though the content of termbases may be identical, it will be displayed in a different way in the different applications. Some tools display the termbase contents in the form of a grid, occupying one of the several panes e.g. DVX; while others,

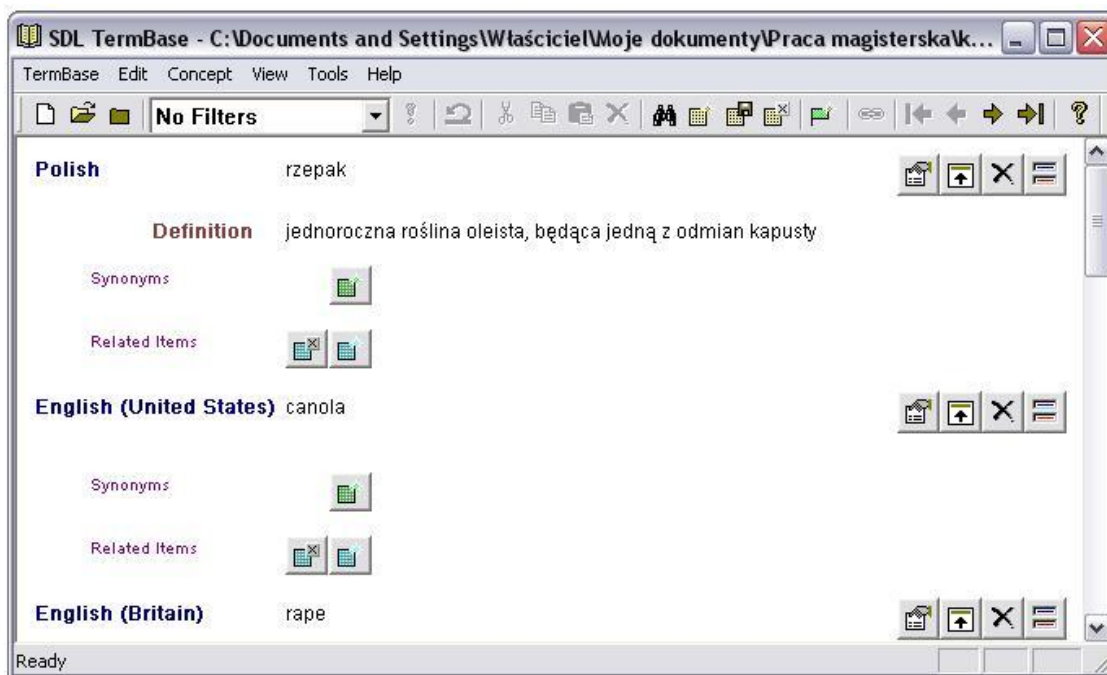


Figure 3. SDLX 2004 – termbase created for the purposes of this thesis.

like Multiterm *iX* and SDLX, display the entries in the form of text files indicating fields with colors, indentations and font sizes. The interface usually allows a number of primitive actions in order to achieve the same effect e.g. a new entry can be created either by clicking a button, selecting the function from a menu or by striking a hotkey. Another feature is the dialog language, i.e. the language in which all the menus, commands and messages are displayed. In most cases several dialog languages are available and can be switched while working with the tool.

Further elements of the interface are the forms of help available to the user. In terminology management tools, usually a user manual or getting started manual are provided with the software or can be downloaded from the manufacturer's webpage. Other forms are tutorials and demos which can best illustrate the usual workflow and basic functionalities of a given tool. Sometimes, also sample files or termbases are provided along the tutorials. Unfortunately, the documentation is rarely available in many languages, thus, it is unavailable to those who do not know English. Furthermore, documentation rarely covers troubleshooting. In order to receive assistance in solving problems the users usually register

in a mailing list. The listmates share the solutions and workarounds they came up with over the years of experience with the tools. This kind of help is extremely important, even though there is usually a possibility to contact the manufacturer and report the problems.

It may also be significant whether the information on the internal workings of the tools – their architecture – is made available. In most cases the codes of a tool are protected by the manufacturers and very little information is given to the users in order to prevent dishonest use of such information by competitors or software piracy.

The way in which the user communicates with the program is by far the most significant characteristic of the interface. Contemporary applications have a wide array of available means of communication. These include typed commands, function keys, traditional menus, pull-down and pop-up menus, dialog boxes, icons, clickable buttons, and wizards. All these elements can be used separately or combined e.g. the same task may be completed by striking a hotkey, selecting the function from a menu or clicking a button.

In most terminology management tools it is possible to manipulate the taskbars, menus, as well as the size and arrangement of the different panes. Also the color-coding of different fields in termbanks and glossaries, as well as translation memories can be customized. It is vital for instance, if the user finds it difficult to read the text from a brightly colored background, which was the case in the earlier versions of *Déjà vu*⁹

3.4 On-screen display

On-screen display is usually definable by the user. Using different techniques in different programs, the users may easily resize and rearrange the taskbars and windows. Also the coloring of the field status indications can be customized. All the changes made by the user in the default display are usually clearly visible or are indicated with a tick on the menus and can be easily reset to default if needed.

In some applications the contents are displayed in the *what you get is what you see* (WYGIWYS) manner, meaning that the users can see the effects of their work in the way it will look in the target format. However, in many CAT tools a translation grid is used instead, which shows both the source and target language together in a tabular form (Assénat-Falcone 2000). In order to see the final effect, though, it is sometimes possible to activate a preview function, also referred to as ‘external view’, which can indirectly display the working file in a final format without the need to conduct the full export procedure.

⁹ In *Déjà vu 3*’s brightly colored background in the translation grid indicated different match status or damaged formatting e.g. formatting changes caused the whole field to be marked with a bright red background.

Most terminology management tools have a default display layout at their disposal and a possibility to modify it and save as custom layouts for later use. Other tools, like Déjà vu X, offer a number of fixed layouts characteristic of particular terminology collections or termbanks.

3.5 Data management

When describing terminology management tools, we should definitely discuss the issue of language support. The list of languages supported by the tool is most certainly the first piece of information considered by a prospective user of the application. Another question is whether all the languages supported by the tool can be used both as source and target languages. In some cases also the support of language varieties is crucial. Moreover, it is sometimes essential that the tool supports bi-directional languages¹⁰ and double-byte character sets (DBCS).

Another question of great importance is whether the terminological entries may only contain textual data. Most terminology management tools are designed to contain textual data exclusively. However, recent innovations enabled the users to store multimedia files including images, audio and video files in Multiterm *iX* termbases.

One other issue of utmost importance is the support of different file formats. In modern translation business it is difficult to imagine working on text files only. Translators are no longer surprised to receive spreadsheet or power point files to be returned translated in the original format. The modern CAT tools usually support a number of data formats including not only the above-mentioned ones but also database files, help contents, a number of DTP formats, java properties, html, sgml, xml, and other. SDLX 2004 also offers a solution for the formats unsupported by the tool itself. It is possible to copy the text to the clipboard and translate the clipboard contents in the program.

One important development is that there are no limits to the number of termbases which can be created by the user. The only limitation can be the size of the hard disc. It is often possible to consult a number of termbases at a time, however the number is not great – about five termbases. Usually also the lookup order can be set up by the user. Furthermore, the number of languages of the termbase can be greater than just one source and one target language. The tools subject to testing within this project offer a possibility of selecting at

¹⁰ Bi-directional languages – typical of the Middle East – are characterized by the bi-directionality of the different elements of the text, e.g. in Arabic the text is directed right to left, while numbers left to right.

least twenty languages for a single termbase. The users may switch the target and source languages freely and sort the terms according to a particular language selected as source.

Terminology management tools now offer a number of statistical data including termbase size, the number of entries and the number of perfect and fuzzy matches. However most statistical analysis is carried out by the translation memory module and therefore should not be included in the description devoted primarily to terminology management modules. What these modules do offer, are basic quality management functionalities, such as warning about duplicate entries, project status, spellchecking¹¹ and terminology validation. In one of the programs tested it is possible to restrict the access to the project or termbase by setting up a password. Another operation which can definitely improve the functioning of the termbase is data compression. Most tools also record the time and name of the user who introduced any changes in the termbase.

Unfortunately none of the tools presented in this thesis creates backup files automatically. Nonetheless, all of them support long filenames, which was not the case with the earlier versions. Years of gaining experience and perfecting the tools subject to study in this project make them stable and reliable. However, problems may occur with each of them and the tools signal them to the user by showing error messages.

3.6 Entry model and structure

There are three possible entry models: free, fixed or quasi-free. The difference between them is primarily in what margin for manipulation is left to the user. In fixed record structure the user can only select one of the available record templates without the possibility to modify it. Using this record structure we may decide to only omit certain fields. The scope of manipulation is greater in the case of quasi-free record structure, which can be defined as a fixed record structure, with one or several user-definable fields. Finally, the free record structure means that the record structure can be defined by the user entirely, including also field naming and arrangement. When designing a free record structure, the users are usually allowed to create picklists of values for the different fields. Most commonly these include subject, customer, project, as well as grammatical information or usage labels. Unfortunately different tools use different field naming conventions, as do large termbanks.

In the case of some terminology collections it is crucial to determine whether there are some limitations as to the length of each field. The tools selected for analysis in this

¹¹ Some tools rely on word processors to conduct spellchecking, other have their own dictionaries against which the spell-check is run.

project theoretically do not impose any such limitations. However, the longer the field, the more difficult it may be to see the content of the whole field or record. Another vital aspect, especially in terms of translation work, is the number of fields for which entry is obligatory in order to create a valid record. The tools discussed here require only the term or lemma. However it is possible to define other fields for which entry is necessary in order to save the record. This operation may contribute to an increased quality of the terminology collection, even though it may significantly lengthen the process of populating the termbase.

All tools under scrutiny offer the possibility to create cross-references across the records. Most often they are created manually and can be found either within a text field or in a separate field. Unfortunately, it is not yet possible to create links to external resources within the terminological records of the tools discussed.

In addition to considering the limitations of the record field length, also the question of the record size should be answered. In none of the tools discussed is there any explicit limitation, however, it should be borne in mind that the larger the records, the slower the data retrieval process may become. It should also be remembered that creating small records enables the storage of much larger data collections in one file without hampering the functioning of the tool. Hence the size of the records should be kept close to the bare minimum in order to retain high speed of information retrieval. Again, there are no limitations to the number of terminological entries, except for hard disc size.

3.7 Retrieval of information

The principal purpose of terminology management tools is to provide quick access to terminological data. Therefore, it is important that they provide a number of efficient search options. Exact match option is very often insufficient. It is especially true for highly inflectional languages such as Polish, where a given term would have to be entered into the termbase in all cases, numbers and genders in order to be found by the tool when working with a text or when automatic term recognition is activated. Therefore, most tools also offer:

- partial match,
- right truncation (omission of the final part of the term in the search string),
- wild card search (replacing an unknown character or characters with an asterisk or a question mark),
- free text search (searching all the text fields of the termbase in order to find the search string),

- fuzzy search (listing all the nearest matches to the search string);
- search via record/translation unit number;
- KWIC (key word in context)
- Boolean operators (operators signifying logical connections AND, OR, NOT);
- relational operators (>, <, =, ≥, ≤, ≠)
- morphological (suffixes, prefixes, inflectional endings)
- search by synonym, cross-reference, internal/external link
- proximity
- subject area
- global search
- restricted (filtered)
- by segments containing a term or phrase (if the translation memory module is active)
- capital vs. small letter
- punctuation and spacing variation
- formatting and mark-up features
- search history
- search log (file recording all search procedures within a given work session)
- browsing (alphabetical, chronological and conceptual)
- access via any data category (e.g. part of speech)
- query language, e.g. SQL

Some applications also allow for a combination of search criteria. Most of them offer the function of global search and replace.

Finally, since most CAT tools are designed as language independent applications, they rarely recognize part of speech of a term, especially if the term has the same form as a noun, verb or adjective.

3.8 System's responses

The system's response if the search criteria are not met is another point worth considering. Usually terminology management tools display a hitlist of near matches, or a 'term not found' message. Some tools also log the unfound terms in a special file, as well as keep a history of searches. For translators working under time pressure, getting the right term right away is extremely important. Usually, they cannot afford delays caused by using a

wrong term found on the hitlist. Therefore, it is vital that the hitlist distinguishes the fuzzy matches from exact matches. In most terminology management tools fuzzy match must be activated manually, so the translator realizes that the hits displayed in the search results may not be appropriate.

Another important question is whether the tool is able to recognize a misspelled term. The logical assumption is that if the tool offers a fuzzy match search option, it should be able to return a misspelled term. The problem is, however, that sometimes a simple typing error may alter the initial letter or letters of the term, thus causing the tool to return absolutely irrelevant matches. This means that the type of the spelling mistake is not without significance.

One more important issue is the tool's response to compound term searches. As has been already said, the tool usually displays the list of nearest matches. With the advent of hybrid systems, combining CAT with MT, some of the tools offer the so-called 'assemble' function, which causes the tool to automatically create its own proposals for rendering the unfound compound term or phrase. Again, the user must be aware of the fact that the output string is not a validated term, but a proposal of the MT module, and as such should be approached with a dose of mistrust and subject to verification.

For the users of inflectional languages it may also be crucial to determine whether a tool returns the canonical form of a word if the search string contains an inflected form. The tools discussed within this project only return forms entered in the termbase. This means that in order to receive a base form of a given term we need to activate fuzzy search, right truncation or wildcard search.

For the users of English it is in turn important to know whether the tool recognizes spelling variants e.g. *colour* vs. *color*; or compound spelling differences e.g. hyphenated vs. non-hyphenated variants.

3.9 Input of information

The input of information into a termbase has a number of aspects. One of the most important ones is whether it is possible to format characters and paragraphs in any way. It is essential in the case of languages where diacritics must be used. Unfortunately, not all fonts offer diacritical characters. Hence, it may sometimes be necessary to change the source text font to another for the target language.

A convenient terminology management tool should also offer a possibility to easily edit both new and existing entries. Some of the simplest editing functions, except for manual

keying in the data, is copying, pasting, dragging and dropping, deleting, redoing, undoing, inserting, and changing the layout, search and replace. These functions are offered by most contemporary applications, including terminology management tools.

3.10 Terminology extraction

Terminology extraction can be defined as an automatic collection of terminology from a mono- or bilingual corpus of texts. Such functionality can highly speed up the completion of pre-translation tasks which include terminology work, since the terminology lists created in the process of automatic extraction can be quickly imported into a termbase. Unfortunately, efficient terminology extraction tools usually do not go along terminology management modules or workbench packages. One exception is the ‘lexicon’ function in *Déjà vu*, which is a very simple, but efficient terminology extraction tool. The manufacturers of the other tools discussed do provide terminology extraction tools, however they need to be purchased separately.

Terminology extraction from a bilingual corpus is probably the most desired function of a term extraction tool. It is even more attractive due to the fact that the bilingual corpus can be created by the users themselves through alignment. Although alignment is more frequently applied for creating translation memories, the corpus created through this process can also be used for terminology extraction. As Feder defines it:

‘Alignment refers to turning two texts into a TM database file and it basically consists in (properly) storing SL segments and the corresponding TL segments as pairs (translation units). For the alignment to be executed properly, first the SL text and its translation must be segmented (i.e. divided into translation segments). Segmentation is also very important when working with a new TM file (i.e. inputting only the SL text and only subsequently producing its translation) [...]’ (Feder 2001:163)

Bearing this definition in mind, we should realize that the alignment function primarily belongs to the translation memory modules, or is offered as a separate tool. Therefore, alignment tools should not be looked at as part of terminology management modules. Hence, they fall out of the scope of the present project.

3.11 Validation/control

Terminology databases more often than not are populated automatically through external termbase import. Automatic import of terminology involves limited control of the data entered during the process. Therefore, some measures of control can be set before import. For instance, using most tools the translator may decide whether to overwrite, merge or import as duplicate the entries of the same term. Automatic terminology validation usually refers to the function which checks the database for duplicate entries. However, when the terminology database is used in the translation memory mode, validation also refers to confirming whether a given source term has always been rendered identically in the target language. The consistency checks are usually run at the end of translation projects and are part of quality assurance functionalities offered by terminology management tools.

Another control measure is spellchecking. It is extremely important to be able to spell-check the contents of a termbase for a number of reasons. First of all a misspelled access term¹² may cause the term not to be retrieved from the termbase, although the translator or terminologist did spend time establishing an equivalent target language term. What is even worse, spelling errors may cause the misspelled term to be inserted in the target language text. The fact that specialist terms usually are not included in the spell-checker's dictionary makes the detection of spelling errors, which are inexcusable, even more difficult and time consuming.

3.12 Exchange of information

The ease of exchange of termbase contents has become a must in the globalizing translation market. Therefore, it is essential that terminology management tools produce termbases whose contents can be exchanged either between users or applications. The easiest form of data exchange is printing. The applications presented in this thesis either support printing directly, e.g. Multiterm *iX* has a 'print this entry' function, or, require a regular export procedure to a text or spreadsheet format which can later be printed (Déjà vu X and SDLX 2004). The fact that the user may decide to export only certain data for printing, e.g. only terms and their translations, or only a pair of languages matters as well.

Nevertheless, the most indispensable feature of modern terminology management tools is the import/export functionality. It allows for a quick conversion of the native formats of a given tool into more universal ones for the ease of exchange between users and

¹² Access term is any search string which can lead the user to a given terminological entry.

applications. The most common import and export formats are text and spreadsheet formats. Recently also the XML format has been gaining significance. Import and export are automatic processes which can save a lot of time and unproductive mouse clicking and typing. However, it should be borne in mind, that there are a number of criteria which should be user-definable, so that the imported data complied with the existing termbase definition and contents. For instance, it must be possible to set the same languages as target and source. Also the in-born limitations of exchange formats must be considered, e.g. the fact that while using MS Excel, we may only use the first sheet, and there must not be any empty columns between the target and source or other field categories.

Another aspect is the record structure. Albeit importing termbase files in the native format does not pose much difficulty, the situation with non-native formats might turn out risky. Not only can the import process be unsuccessful, but it may also bear serious consequences for the termbase which was to be populated with a new terminology collection. Therefore, data designed for import should be well prepared and termbase parameters set very carefully. Also back-ups of all files involved in the export/import process should be saved to other directories on another disc or CD, if possible.

Export, on the other hand is a process which sends the contents of the termbase into an external format. This process should also be preceded by a careful definition of the parameters and properties of the export files. The users should take into account the purpose of the procedure and the destined use of the exported termbase. To help decide what data and in what form should be exported or imported the software engineers designed export and import wizards which guide the user through subsequent steps, prompting for selection of necessary criteria, file names and locations.

As has been mentioned, the most widespread termbase exchange formats include text formats (most common are delimited text files such as comma separated values (*.csv) or tab delimited text files, spreadsheet formats (MS Excel) and XML. A fairly new exchange format - TBX – Terminology Base Exchange Format – is an XML- based terminology exchange format comparable with the TMX created in order to enable seamless exchange of translation memories between different applications. TBX is supposed to provide the same ease of exchange of terminological data between applications and users. According to Localization Industry Standards Association (LISA), TBX enables exchange of data between termbases with different data models (Melby 2001).

Finally, terminology management tools may support the native formats of other CAT tools e.g. DVX supports Trados format and SDLX supports DV and Trados-derived formats.

3.13 Interaction with other applications

In the translation and localization business nowadays, it is impossible to rely on stand-alone terminology management tools only. Therefore, the tools discussed in this thesis usually can be integrated with other applications in order to create more complex translation environments. The most frequent combinations are:

- Termbase module and a word processor;
- Termbase module and a translation memory (depending on the tool this environment may involve the word processor as well);
- Termbase module and a machine translation system;
- Termbase module and a concordancer i.e. a tool which analyzes source texts to deliver ‘an alphabetical list of all the words in a text shown together with their context and references to line in source text’ (Khurshid 1994);
- Termbase module and a word frequency program;
- Termbase module and a speech recognition system¹³.

The integration of terminology management tools with the above-mentioned applications usually enhances the applicability of the termbase module itself, especially that these tools are nowadays primarily intended for use as integral parts of larger translation environments.

If it is possible to combine the termbase with a word processor, many functions of the tool are available directly from the word processor interface. Usually a new toolbar or menu appears next to the regular WP toolbar after the selection of an appropriate template from the program installation folder. The integration allows for an automatic lookup of terminology found in the source file. Entering new terms into the termbase is usually possible. Moreover, the procedure for entry creation from within a word processor is usually reduced to the bare minimum, i.e. entering the source and target language term. Other functionalities available in this environment include direct copying and pasting both from WP into the termbase and in the opposite direction, as well as automatic replacement of the terms found. Since new information may be found while working with a text, it should be possible to modify the

¹³ Speech recognition systems are in this context understood as tools which allow the user to dictate the text instead of typing it. According to the members of CAT software mailing lists, the combination of these systems with CAT tools highly increases the translation productivity. The translation dictated instead of typed has also this advantage that it is usually of better quality, as the user can hear the translated segment as a whole. Consequently the translator produces utterances which are more stylistically correct. It is important to know that using a speech recognition system is not mutually exclusive with using other input devices (keyboard, mouse, trackball, etc.), so the functions activated by a single keystroke, instead of uttering several words, are still active.

existing entries to enter the new findings in the WP mode. Another important functionality, unfortunately not implemented in all the applications discussed, is that several data collections should be available for reference at the same time, without the need to merge them into a single termbase. Finally, just as in the case of the stand-alone termbase modules, the logging of the unfound terms is a desirable functionality.

Another possible step in termbase module integration is the integration with translation memories, i.e. creating the so-called workbenches. As it has been mentioned, combining the terminology management tools with other modules, in particular translation memories enables certain functions which are unavailable in the standalone mode. Therefore, for instance connecting Trados' Translator's Workbench with Multiterm *iX* enables automatic term recognition. Another Workbench function allows for the automatic replacement of the terms found in the termbase. Some tools allow the insertion of a single term identified at a time, while others replace all the terms found in one operation. All the translation memory tools are tailored to the specific workflow translators usually follow. Therefore, it is possible to add term pairs to the termbase while actually translating in the workbench. This procedure is usually of a summary nature in contrast to the process applied when manually creating a new termbase from scratch. Most rapid entry options create terminological entries containing only the target and source terms. However time-saving this may seem, this may be a shortcoming for those who store all their terminology in a single termbase, but have a number of customers who use different corporate language. In this case, it is indispensable to enter the customer's name in the entry, to make it useful for the future translation projects.

Another issue worth considering is the number of termbases which can be consulted at a time. Some workbench tools allow for consulting multiple termbases, others do not. Therefore, it is crucial to choose the tool which best suits our preferred style of working, i.e. if we have separate termbases for each client, it is better to be able to consult several of them at a time. However there are simple solutions to the limitation of CAT tools in this respect. It is usually possible to merge different termbases by way of import, or export. Another significant aspect is the way the search results are displayed if a number of data collections were consulted. In order to make appropriate terminology choices, it should be possible to identify the source termbase of a given term in the hitlist. Some tools solve this problem by opening separate hitlist windows for each termbase, others produce a single hitlist but arrange the terms found by termbase in the order defined by the user.

Other issues, also present in the integration with word processors include the ability to log the terms unfound and all unsuccessful queries, as well as the possibility to modify the termbase contents. Both these aspects are important for the smooth flow of translation work in any configuration of the terminology management systems.

With the increasing tendency to create more complex and automated translation environments terminology management collections are more frequently enabled for machine translation (MT) system integration. The use of terminology collections in TM systems requires a specific entry structure and usually more linguistic information than is normally necessary for human translators who can simply rely on their linguistic competence and intuition. Nowadays there is a tendency to create such terminology management tools which can be used both for translation memory and machine translation systems. Also some developers of translation memory software provide MT systems as well, e.g. SDL International (SDLX), or Star (Transit).

Automatic terminology extraction is another significant development in translation technology, therefore some forms of term extraction tools are either combined with the latest releases of CAT tools, e.g. Atril (DVX), or offered as modules which need to be purchased separately e.g. Trados (ExtraTerm) and SDL International (SDLX).

Word frequency programs, alignment tools and concordancers are other examples of tools with which it is possible to combine terminology management tools. However, they frequently constitute part of the workbench environment and are recognized as separate functions rather than tools. Word frequency programs or functionalities allow translators to establish whether a given project is suitable for the use of CAT tools in that it analyses the repeatability of terminology and certain phrases in the source text. Also, they allow for the better calculation of the actual terminology and translation work that needs to be done, providing a thorough basis for the correct and fair billing of the client. Alignment tools in turn allow for efficient semi-automatic terminology extraction. Some versions of CAT tools (e.g. DV3) allow for automatic population of the termbase with the term pairs identified as a result of alignment. Finally, concordancing programs allow for the indexing, key word identification and analysis, identification of phrases and fixed expressions, and comparison between the different usages of the terms. Just as in the case of alignment tools, which are frequently provided with translation memory modules, many workbenches have the concordance search functionality e.g. Trados (Translator's Workbench). Another solution which to a large extent addresses the needs solved by the above-mentioned tools is the Lexicon function in DVX. Lexicon is created separately for each translation project. Once

the import of files is completed, the project lexicon should be created, which lists all the words of the project listed separately and in the context of the number of words set by the user. Thanks to this functionality the frequency count of each term and phrase identified is provided. In this way, the translator receives a concordance and word frequency analysis. Batch deleting the unnecessary phrases, and translating the remaining core terminology of the project, the user may either leave the lexicon for reference as a separate source, or feed its contents to the termbase immediately.

Having considered the vast possibilities of combining the terminology management tools with other applications in order to support translators in as many aspects of their work as possible, we should realize that those integrated applications are indeed powerful.

3.14 Fonts and character sets

Several words should be devoted to the fonts and character sets in the context of terminology management applications. Most terminology management tools support the default Windows fonts and character sets. Even though problems with certain special characters are reported by the users, the support of diacritical characters and DBCS is increasingly satisfactory, given the fact that the latest releases of the CAT tools are Unicode compliant. It is important to determine whether the special characters and fonts supported by terminology management tools are also supported by the other modules, and functionalities, for instance whether it is possible to transfer the special characters among different application windows e.g. between the context and definition field, or between the WP and the termbase. Another significant aspect is whether the specific language support means the support of this language's sorting order.

3.15 Maintenance operations

Termbase maintenance operations include saving the database, updating the index, termbase compression and repair, generation of backup files. Updating the termbase index is an operation necessary for terminology retrieval to be efficient and up-to-date. In the earlier generation of terminology management programs this had to be done manually, however in the latest releases e.g. Trados Multiterm Workstation, this task is performed automatically. Termbase compression is indispensable to speed up the search process and the efficiency of term retrieval. The applications discussed in this thesis can also retrieve or repair corrupt termbases, protecting in this way the valuable terminology collections from destruction. Saving the backup files should be a regular procedure of any translator working with CAT

tools. It may turn out a useful routine operation not only in the case of application or operating system failure or irreversible damage of a termbase, but also when unwanted changes have been made to the termbase content, and the valid entries have been altered.

3.16 Commercial aspects

Finally, the commercial aspects of terminology management tools are the decisive factor as far as the presence of the applications on the translation market is concerned. These aspects include the developer and distributor of the tool and strongly affect the tool's position and perception. Another crucial issue is the pricing, which unfortunately is nowadays the key deterrent for the freelance translators. However other aspects, such as the frequency of issuing updates, and upgrades, tool availability on the domestic market, the number of users, also the renowned ones, determine the popularity of a given terminology management tool among translators.

4. TERMINOLOGY MANAGEMET TOOLS – DISADVANTAGES AND PROBLEMS COMPARED WITH THE BENEFITS

As mentioned, in the introduction to this chapter, this section shall be devoted to presenting the author's remarks concerning the advantages and disadvantages of the sample applications tested in this project. For the sake of convenience the comments are broken down to the separate tools and split into advantages and disadvantages.

4.1. MULTITERM

4.1.1. ADVANTAGES

One of the greatest advantages of Trados Multiterm *iX* is its self-explanatory Windows-based interface. The Multiterm *iX* Wizards guide the users through all the procedures step by step. The user-friendly design of the tool is definitely encouraging for a novice translator or a person who only starts working with it. It is also worth mentioning that Multiterm *iX* is rather fast to set up and begin the actual work.

One other undeniable advantage is that the XML format support has been introduced into the application. Freelance translators do not have access to network termbases, so it is extremely important that their terminology can now be stored in the format which facilitates the ease and speed of data exchange.

According to Nogueira: 'Trados works and works well, mainly in the translation of Microsoft Word documents, which is what many of us translate all the time' (Nogueira 2002). Effective translation of texts is what the majority of translators aim at, and a good tool able to cooperate seamlessly with the most frequent file formats is hard to overestimate.

Finally, MultiTerm fulfills its role as a source of consistent terminology. The fuzzy matching option allows for the recognition of terms even when they are used in a different grammatical case than in the termbase. This is of utmost importance when translating from inflectional languages such as Polish.

4.1.2. DISADVANTAGES

Having browsed through the 325 pages of the MultiTerm User's Manual any translator has the right to feel overwhelmed. The flexibility of the tool, the staggering number of available settings, and customizable functions in every operation might turn disadvantageous. One of the professional users of CAT programs aptly expresses the frustration that accompanies translators when using heavily customizable applications:

'DV and Trados offer you a lot of various windows and functions that may be useful to computer wizards but scare away the majority of translators. They are basically meant for translation managers and translators who are also very advanced PC users. Besides, all these sophisticated functions are very seldom used, if ever.' (Gerasimov 2001)

Another comment seems to express a similar opinion:

'MultiTerm, the terminology management utility, is a Rube Goldbergian piece of software that requires no less than 12 steps to add a new term to a glossary, whereas most competing programs require only three.' (Nogueira 2002)

This statement referred to the 5.x versions of Multiterm. The new Multiterm *iX* has to a large extent addressed the problem of too much complexity and is now much more intuitive and user friendly.

Another development which can be considered an inconvenience is that when using Mutlitem *iX* in combination with other workbench modules or word processors – which is the most frequent configuration – only one termbase can be consulted from within the integrated environment.

4.2. DÉJÀ VU

4.2.1. ADVANTAGES:

Among the many advantages of the way Déjà vu X handles terminology, is the function of the Lexicon, i.e. a project-specific bilingual/multilingual list of terms and phrases. Even though the fact that only one termbase can be associated with a given translation project may seem an inconvenient limitation, the project-specific lexicon makes up for this in a number of ways. On the one hand, the project-specific terminology and phraseology contained in the lexicon is ready at hand during translation. On the other, the new terminology does not have to be sent into the termbase until the project has been completed and all terms validated. Hence, the main termbase is not fed with invalid terms and the time which would have to be spent on finding and correcting the erroneous contents of a large terminology collection is saved.

Another undisputable advantage of this tool is that whatever file type a translator works on, Déjà Vu presents them with a uniform interface, displaying only translatable text and protecting the tags. This assures comfort in performing translation tasks without having to worry about overwriting the formatting and layout information.

Numerous spell and consistency checking functions help the user ensure the publishing quality of the output. One other noteworthy feature of Déjà Vu is that it is very easy to add or modify the contents of the database. Thanks to the ease with which the lexicon, once edited and translated, can be entered in the terminology database, the time spent on preparing a translation project can also be reduced to a large extent.

Finally, the on-line support of Déjà Vu deserves a special mention in this section. The supporting team of Atril is extremely helpful. The response to any technical problem reported can be expected within a few hours, including weekends, which is not commonplace on the market. The users are invited to join a mailing list where the members share their problems and try to work out the solutions together. Even passive membership of the mailing list can turn advantageous for the beginners, as they may learn from the experienced translators and users with an impressive expertise concerning the program and its numerous versions.

All these features, and many others which had to be omitted due to the inherent limitations of the theses of this type, contribute significantly to the improvement of speed and efficiency of translators without affecting the quality of their output.

4.2.2. DISADVANTAGES

Being convinced of the many advantages the tool has, we should not turn a blind eye to the imperfections it has as well. As has been mentioned above, Déjà Vu as a whole discourages novices in translation and inexperienced computer users with its complexity (Gerasimov 2001). We should bear in mind the sad truth that still not many of the linguists, who most frequently become translators, reach advanced levels of computer literacy. Moreover, even a careful study of the 534-page-long User's Guide usually turns out insufficient for a beginner.

It is not perhaps surprising that using DVX with the most popular languages causes fewer problems than in the case of minority languages. The users who work with minority languages, e.g. Japanese, report display problems. Also diacritical letters e.g. the Polish 'ż' turn out to cause some problems. A slightly different issue related to display is the fact that in AutoSearch mode the application does not display term definitions, which may be a serious drawback if the termbase contains multiple translations of the same terms (e.g. client-specific terminology).

Another difficulty is the reliability of the source file analysis. Some discrepancies have been reported between the word count performed on the same document by DVX and Trados Workbench, however it is difficult to attribute the problem to either application.

One issue which can certainly be considered a disadvantage is that the once-existing function of the automatic feeding of the termbase with the term pairs collected during alignment has been removed in the DVX version.¹⁴

The Déjà Vu listmates share the problems they encounter while installing the new upgrades or the new version of the program – DVX. The diversity of complaints regarding DVX ranges from installation problems to slow conversion of older DV databases, to file import problems. Some of the users also report OS¹⁵ compatibility problems. This diversity is highly indicative of the complexity of the software.

¹⁴ Atril's support team promised to re-introduce this function in one of the upcoming builds.

¹⁵ OS – operating system

4.3. SDLX

4.3.1. ADVANTAGES

Finally, several comments will be devoted to the SDLX™ 2004 translation suite. One undoubted advantage of this tool is that there is a similar user interface in all modules. This increases the speed of getting familiar with the tool (Waßmer 2003).

Also the relative simplicity of SDLX™ 2004 as compared to the other ones is probably one of its greatest advantages. The color coding keeps the users informed about the status (match level) of each cell in the project. Database navigation facilities (TermBase minibrowser, the saving of search filter and settings, internal hyperlinks) contribute greatly to the ease and efficiency of search procedures.

The extensive list of supported file types (including among others Trados export format) is also an advantage for the users, who can focus on translation tasks without having to learn how to use all the file formats. It also rids the users of the need to purchase all the expensive software.

The list of languages supported by the tool, which is now fully Unicode compliant, is impressive and includes BDCS and bi-directional languages.

4.3.2 DISADVANTAGES

One of the disadvantages of the tool is the inconvenience of the new termbase field definition e.g. no multiple selections are allowed when selecting values from picklists. Another problem is the fact that no graphic or other multimedia files can be entered in the termbase fields.

Even though MS Word is probably the most popular computer software, the fact that it is required in order to translate the *.doc files in SDLX™2004 may be a problem to those who rely on other word processors.

Also the fact that there are no external dictionaries can be associated may be perceived as a disadvantage. Another difficulty is that automatic search cannot be performed simultaneously in translation memory and terminology database.

5. CONCLUSION

Following the description of a workbench-supported translation process, this chapter contained a detailed presentation of the functionalities which terminology management tools

offer, as well as some aspects influencing their usefulness and popularity among translators. The last section contained remarks on what are some of the most significant advantages and disadvantages of the sample tools. However explanatory such a description may be, it remains general and thus of little significance for translators, linguists and any parties interested in the subject, if not supported by specific examples. In a nutshell, at this stage the readers are not given enough information to determine whether these tools may be of assistance in their particular working conditions. Also the scientific implications of this type of presentation are scarce. Hence, it is necessary to immerse in the analysis more deeply, testing the tools separately against objective criteria.