[*Translating and the Computer 28*, November 2006 (London: Aslib, 2006)]

GOING GLOBAL WITH TextBase TM Translation Memory

The Best of both Worlds!

François Tardif
Customer Service - Europe
MultiCorpora International
+32 2 420 9600
ftardif@multicorpora.com

Anne Laugesen
Translator
Global Denmark - International Communication Consultancy
+45 33 86 29 41
anne@global-denmark.dk

MultiCorpora International
Rue de STASSART 117
B-1050 Brussels, Belgium
www.multicorpora.com

Revised: Oct-25-05

Introduction

Driven by fast-paced global competition where the time-to-market of new products, services and communications into multiple languages and cultures is mission-critical, organizations are increasingly demanding translation services that provide faster turnaround while maintaining the highest level of quality. A key driver behind the need for speed and quality is the ongoing explosion of web-based content and the related expectations of content freshness and quality. Operating in a competitive and typically fixed-price environment, translation service providers need to respond with significant gains in translator productivity while continuously improving translation quality.

For many years now, Computer-Aided Translation (CAT) tools have held the promise of productivity and quality gains for translators. Despite considerable hype by the early software vendors, the first generation of CAT tools has failed to deliver. The up-front investment to populate multilingual terminology and translation memories has proven prohibitive for many service providers and the actual productivity gains realized have been insignificant except for a few, very specific types of content.

Simply stated, translators are skilled writers who must take information in one language and communicate it in a different language while maintaining the precise meaning, style and tone of the original communication. Achieving these objectives requires much creativity and linguistic skill - skills that computers lack. While machines might be capable of generating approximate language translations for the purposes of indicating the gist of a written passage, they will likely never replace human translators for communications where accuracy and quality are important. With this firmly in mind, a different computer-based approach to helping translators has been developed and the industry is finally beginning to realize significant gains in translation productivity and quality.

This paper explores the promise of translation memory and the inherent limitations of traditional CAT tools. It then uses this as the foundation for introducing a different approach: the full-text multilingual TextBase Translation Memory.

The Promise of Translation Memory

Conventional wisdom holds that there are few (some say no) original ideas or thoughts, just reflections or reinventions of the previous ideas of others. In translation work, this holds very true. Research clearly shows that complete sentences rarely recur from one document to the next but repeating smaller expressions (mainly five words or less) are very frequent. For many types of documents, recurring expressions account for over 50% of the words in the document. For technical documents, the level of coverage can exceed 75%. Even in the least-repetitive types of content, such as administrative communications and parliamentary debates, research shows that the level of expression repetition exceeds 25% of the words in a given text. In all cases, the majority of the repetition comes from expression of 5 words or less. Complete sentence repetition, however, is almost non-existent in all types of content except technical manuals for products that belong to a large family of similar products.

This high degree of repetition of expressions means that a large fraction of any new project has been previously translated many times. These previous expression translations exist in many accessible locations: in previous projects translated by the same translator, the work of other translators in the same organization, the work of other translators working for the same client, and publicly available translation work. The Internet alone provides convenient access to multilingual web sites that contain billions of words of translated text covering every topic. For example, a translator working on a project in the health field can reference hundreds of thousands of high quality translated expressions from the World Health Organization (WHO) tri-lingual web site.

Of the many previous translations for a given expression, some will not apply in a specific context because the terminology usage, style, and/or tone are inappropriate. However, many high-quality examples that fit the current context will also exist. If these previous translations of expressions could be effectively recycled and used as context-sensitive translation references, enormous gains in translator productivity and translation quality could be realized for all types of content.

Traditional CAT Tools Have Failed to Deliver

Almost 15 years ago, Translation Memory (TM) tools began to appear on the market. The original concept was simple and has not changed significantly since then: human translator productivity can be increased and the consistency of translations can be improved if previously translated sentences can be stored in a database for later retrieval. This core assumption has proven to be flawed in several ways that we will explore later.

TM-based CAT tools have founded limited application. The only niche where they have gained real traction is the translation of technical product documentation, such as technical specifications, operating manuals, and maintenance and support documentation.

Despite considerable hype by the software vendors involved, TM-based CAT tools have found limited application. The only niche where they have gained a foothold is the translation of technical product documentation (technical specifications, operating manuals, maintenance and support documentation, etc.). Moderate success has been attained when products share many characteristics (for example, multiple models of photocopiers within a product family) and when documentation requires frequent updating. This type of documentation is characterized by relatively frequent recurrences of whole sentences.

Even in this niche, however, there is a considerable gap between the productivity gains hyped by TM software vendors and the actual gains realized by translators. Vendors typically claim 75% to 100% or greater gains in productivity while independent reviewers cite gains ranging from 15% to 30% under ideal circumstances. The reason for the discrepancy is that CAT tools vendors use an unrealistic *theoretical* model to compute productivity gains rather than *actual* translator experience. Vendors typically use an example where a product update document contains 65% of repeated whole sentences. They then compare the time to translate the entire document manually with the time to translate the new 35% only, assuming that the TM tool takes care of the repeated 65%. This is simply nonsense. The reality is that much of the repetition in that type of document is in large sections of text, possibly entire chapters. Translators would not work through the entire document and re-translate large sections that have not changed. The project manager or translator would first identify the changed sections based on guidance from the author or using the standard "Document Compare" feature available in most word processors. They would then focus their translation efforts only on those new sections.

Research indicates that existing CAT technology is appropriate for less than 5% of all documents that require translation.

Also, any actual productivity gains do not factor in the considerable effort to build the initial TM database before the actual translation work can begin - which for many projects completely offsets any productivity gain. To make matters worse, clients typically demand lower translation price rates when a TM is used, thereby wiping out any economic benefit to the translator.

Outside of the niche of repetitive technical documents, the productivity gains from TM-based tools have proven to be insignificant since repetition of whole sentences is extremely rare in most types of content. In fact, research indicates that existing CAT technology is appropriate for less than 5% of all documents that require translation.

The actual number of translators actively using TM-based tools is probably closer to the 5% mark

Additional proof of the lack of success of TM tools is the limited adoption rates and the fragmentation of the CAT-tools market. Surveys of translators and the public sales claims of the leading vendors indicate that, at best, 15% of translators have purchased licenses for CAT tools. Considering 1) that many of those early adopters have purchased more than one tool set (searching for the promised benefits) and 2) that there are many accounts of translators who have invested in these tools but abandoned their use (creating a considerable "shelf-ware" syndrome in this market), the actual number of translators actively using TM-based tools is probably closer to the 5% mark

Looking at the leading CAT-tools vendors themselves, most operate their tools software businesses as a sideline to their translation services businesses and the combined annual software sales of all CAT tools remains very small. None of these factors indicate a very successful commercialization of technology.

To better understand the lack of success of these tools, let us take a closer look at how they work and their inherent limitations.

Translation Memory Systems Overview

A typical TM-based CAT tool consists of the TM database, alignment tools to help build the initial database, a separate database for terminology management and a “workbench” environment for actually editing translation projects and incorporating the TM and terminology resources. In order to better understand the inherent limitations, let us first look at some of the key components of these solutions.

Pre-Translation

Pre-Translation is the fundamental value proposition of TM systems. It compares each sentence in the new project to the sentences in the TM database and, if an exact match is found, it automatically retrieves the corresponding (aligned) translation and inserts it into the target document. This automatic substitution is usually referred to as “pre-translation” because the human translator then fills in the missing pieces of the translation to complete the project.

Fuzzy Matching

When searching for matching sentences, most TM systems are capable of identifying exact matches and “fuzzy matches” - matches where there are minor differences between the two sentences. In the pre-translation process, fuzzy matches must be reviewed, edited and accepted by the translator. Since exact sentences rarely recur, fuzzy-matching technology has been the subject of much R&D investment and marketing hype.

Some TM-tools vendors are taking this concept even further and applying machine translation and/or searching words and expressions in terminology databases in order to automatically convert fuzzy matches into more exact matches. The results make for great marketing hype but the results have been fairly disappointing - computer applications simply do not possess the intelligence and creativity required to compose high-quality translations, where the standards for meaning, style and tone are very rigorous.

Perhaps a different perspective should be taken on how to better support human translators, rather than trying to patch over the shortcomings of sentence-level TM by attempting to further automate the creative human component of the translation process.

Text Segmentation

Translation Memory vendors use the term “segments” to refer to the text strings stored in their databases. All solutions use similar, well-known algorithms to break up legacy documents into these segments when creating the initial TM database. The same algorithms are applied to new translation projects to break it into segments that can be compared to the entries in the TM database.

For the main body text in a document (where almost all of the translation effort is required) TM systems are limited to dealing primarily with whole sentences, which rarely recur.

These text-segmentation algorithms use punctuation and document-formatting information to extract the individual segments. What gets extracted includes: whole sentences, titles, headings, captions and table entries - all of which can be easily identified from punctuation and formatting. The most sophisticated systems are also able to accurately separate compound sentences by finding, for example, a known conjunctive adverb like "however" that follows a semicolon.

Sloppy use of the term “segments”, however, often results in misleading claims of repetition in documents. An automated analysis (something many tools provide) of a document might indicate that 50% of the segments in the document recur. The

impression is that 50% of the sentences in the body of the text recur. In fact, the repetition of whole sentences might be zero but the segment repetition count is skewed by highly repetitious table text entries, headings, and even footer information such as “page 1”, “page 2”, and so on. One of the most popular TM tools even counts blank lines with carriage returns as a repeating segment within a document. This common segmentation approach means that for the main body text in a document (where almost all of the translation effort is required), TM systems are limited to dealing primarily with whole sentences, which rarely recur; they cannot deal effectively with individual words or sub-sentence expressions - which is where most of the potential reuse occurs.

Translation Alignment

Even for the 5% of all documents where a database of isolated, complete sentences and their corresponding legacy translations is useful, a database of significant size needs to be built before it can be of value to a translator. Most TM systems include some capability to help extract sentences from legacy documents and to align those sentences with their corresponding translation from the previously translated sister language document, thus creating the TM database. Since the TM database is used for automatic pre-translation of new documents, all sentences in it must be perfectly linked to the correct translation.

Alignment tools, often called “auto-alignment” by vendors, do an approximate job that must be reviewed, adjusted if necessary, and approved by a human translator before it is committed to the TM database - there is simply no such thing as fully machine-automated alignment. While some vendors provide graphical alignment aids to streamline the manual process, it remains extremely labor-intensive and time-consuming. Users of these tools report that about 3,000 words of legacy text can be processed into a reliable TM database in an hour. Most organizations would like to be able to leverage large sets of legacy documents, containing between several hundred thousand words and several million words of legacy text. The effort to build a complete TM in that size range is many weeks, in some cases months, of translator effort. Translators simply cannot afford to devote that amount of time to non-project (i.e., non-revenue generating) activity.

To build a complete TM of that size takes many weeks, in some cases months, of translator effort. Translators simply cannot afford to devote that amount of time to non-project (non-revenue generating) activity.

To make matters worse, most of the effort spent aligning legacy text is wasted since only a small fraction of the sentences in a TM will ever be recycled for new projects. Even in the most repetitive niche applications, less than 15% of a TM database might ever be accessed. The entire legacy text must be accurately aligned, however, because it is impossible to predict which 15% will eventually be needed.

A TM database can also be grown slowly over time as human translation work on new projects is completed. Translators using the “translation workbench” environment of a TM tool can commit each sentence in a project and the completed translation to the TM database as they work through the document. By creating, reviewing and approving the sentence translation as they work, they are guaranteeing the quality of the alignment. The catch-22 is that it takes a long time before the database becomes large enough to offer significant benefit back to the translator. At an average translation rate of 2,000 words per day, it could take up to a year to create a large TM database. In the meantime, the translator must accept the use of a new tool that offers no immediate benefit - something that runs counter to human nature, which is characterized by impatience and the desire for immediate gratification. This is one factor that has led to a considerable amount of TM software licenses becoming so called "shelf-ware".

It could take up to a year to create a large TM database.

Integrated Terminology Repositories

Most TM systems include an integrated terminology database to capture and manage special terms specific to an industry or topic area. These databases typically include: terms; their monolingual synonyms; corresponding translations in multiple languages; user-defined information about the term such as its origin, usage control notes, and definitions; and other administrative information. These are important systems that encourage the consistent use of special terms.

Most TM systems also support “pre-translation” from terminology databases, whereby the source text of a new project is compared to the terms in the database. Where matches are found, the corresponding term translation is retrieved from the database and inserted into the new project. Assuming the terminology database is sufficiently large, this is significantly more valuable that the whole sentence matching attempted with the TM database.

Similar to TM sentence databases, terminology repositories are very time-consuming for terminologists to build. Consequently, they are usually restricted to the most important and most domain-specific terms in a given topic area. While they are a valuable resource for translators, terminology databases provide limited usage context and they cover only a tiny fraction of all recurring expression translations that exist.

Concordance Search

A concordance search capability allows translators to search for a specific word or expression in a TM database and retrieve all of the sentences where that word or expression occurs. This manually executed type of search is very slow in most TM systems, since they were optimized to match whole sentences. Many TM users report that they rarely use the concordance search features because of the very slow search speed.

This manually executed type of search is very slow in most TM systems, since they were optimized to match whole sentences.

If the translator has the patience to wait, the search results might provide some assistance; however, because the concordance passage is limited to isolated sentences that exist in the TM database, there is often insufficient context to provide guidance on the applicability of the found result.

Peripheral Functionality

Having taken the core TM technology as far as they can, TM-tools vendors are increasingly investing their R&D efforts in the solution periphery: improving word processing features in their proprietary editing environments, additional project management features for large agencies, and web content globalization workflow. Many of these features provide some level of value to the overall process but none directly addresses the core issues of translator productivity and translation quality.

Problems with Translation Memory

The previous overview of TM technology alluded to several shortcomings. In summary, **the three biggest limitations of TM are:**

1. **Dependence on whole sentence repetition**
2. **Loss of context**
3. **Building a TM database is prohibitively labor-intensive**

These three technical shortcomings significantly limit the positive impact TM systems have on productivity and quality - in some cases actually degrading both productivity and quality. To make matters worse, many translation clients (companies or large agencies who subcontract) insist on lower translation price rates when a TM is provided - offsetting the economic benefit of any productivity gains.

Dependence on Whole Sentence Repetition

While whole sentences almost never repeat, recurring expressions less than 5 words long account for the majority of the text volume in all types of content.

As we have already explored, TM systems are essentially limited to exploiting whole sentence repetition in previous translations. Research into the characteristics of different types of documents in different domains (industries, topic areas, etc.) clearly shows that whole sentences rarely repeat. The only exception to this are certain highly structured technical documents, mainly related to complex product families.

Overall, significant sentence repetition occurs in less than 5% of all documents that are translated. TM systems are therefore largely useless for administrative memos, speech or meeting transcripts, marketing content, and most types of corporate studies and reports. The globalization of corporate web sites is another area where the vast majority of the content is not repetitive at the sentence level. While whole sentences almost never repeat, recurring expressions less than 5 words long account for the majority of the text volume in all types of content. Since TM systems do not effectively address short recurring expressions, they are unable to significantly boost productivity.

Loss of Context

Solving difficult translation problems is time-consuming due to the need for manual look-up of style and usage references, particularly for junior translators or translators new to a domain or particular client. Usage and style context is critical to making sound translation decisions.

Building a sufficiently large TM database takes weeks or months of effort - making it prohibitively time consuming.

Since TM systems maintain a database of isolated sentences, they lose the surrounding context within which the original sentence was used. The lack of style and usage context results in additional time-consuming translation review and editorial rework because translations built from isolated sentences are more likely to contain inconsistencies or errors. Further, by automatically “pre-translating”, TM systems blindly reuse sentences that might not fit the context of the new project, resulting in poor-quality translations. To improve quality, the translator and others must spend extra time and effort to review, edit and correct - resulting in lost productivity.

Building a TM Database is Prohibitively Labor-intensive

The millions of words of potentially valuable previous translations that exist around any given client or topic simply can never be fully exploited by the TM approach.

As discussed previously, building a TM database is a tedious labor-intensive process of ensuring that 100% of the sentences in legacy documents are perfectly aligned with their corresponding translations. With the best alignment tools available, the process of creating a TM takes about 3,000 words per hour of translator time. Some alignment tools take up to twice that time. At those rates, building a sufficiently large TM database takes weeks or months of effort - making it prohibitively time-consuming in the competitive, low-margin business of translation services.

The millions of words of potentially valuable previous translations (internal and external sources) that exist around any given client or topic simply can never, in any practical sense, be fully exploited by the TM approach.

Similarly, building large formal multilingual terminology banks (or responding to a translator's specific request for terminology clarification) is an extremely time-consuming process, because terminologists lack the context-sensitive reference tools to rapidly research and create terminology records.

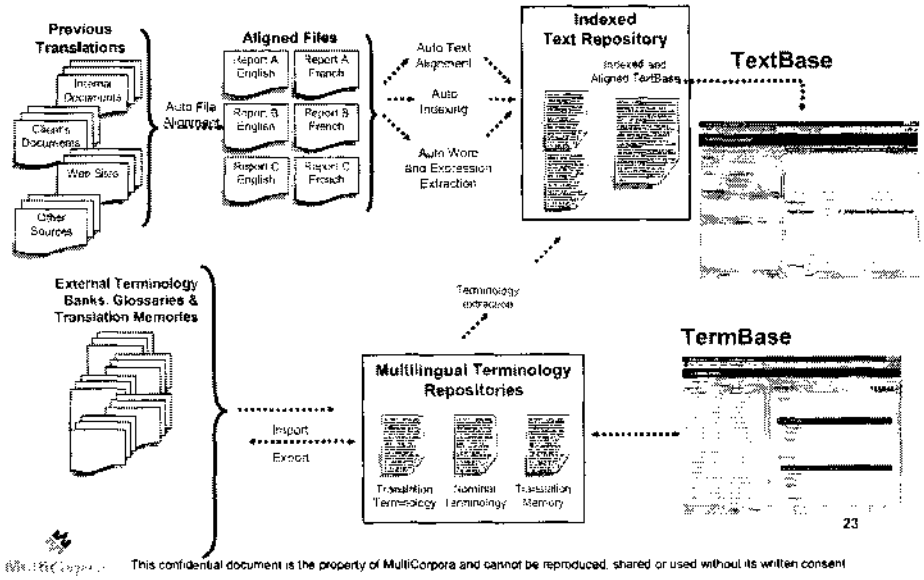
A New Perspective: The TextBase TM Approach

Translation Memory tools were born and developed to address a very real problem: how to leverage previous translation efforts to realize significant gains in translator productivity and translation quality. However, taking the different perspective of focusing on how to better support human translators rather than how to gradually replace them, a fundamentally different - and in many ways more elegant - approach to the problem emerges.

Considering the shortcomings of traditional TM technology, the challenge is to provide a solution that:

1. **Enables the rapid creation of a vast reference pool of previous translation efforts**
2. **Provides complete usage and style context for previous translations**
3. **Effectively recycles translated expressions of any length, not just whole sentences.**

MultiTrans harnesses your existing linguistic assets



The Full-Text Multilingual TextBase TM: Instant Access to Millions of Words of High-Quality Translations

A TextBase TM of millions of words can be built in less than an hour and be ready for immediate use by translators. A comparable size TM database would take years to build.

A searchable, full-text TextBase TM provides a simple, effective solution to these challenges. Rather than tediously building a database that contains isolated whole sentences and their previous translations, the TextBase TM approach takes a vast collection of legacy documents and their previously translated sister documents and, using advanced search engine techniques, rapidly indexes all of the text. It also uses algorithms, similar to those used by TM tools, to align the translated text with the source text. A very large searchable TextBase TM of previous translations can be built very rapidly - at a rate of approximately 50,000 words per minute on a low-end computer. A TextBase TM of millions of words can be built in less than an hour and be ready for immediate use by translators. A TM database of comparable size would take years to build.

The potential benefit to translators of being able to easily reference web content is enormous.

With this approach, it is possible for a translator to build a TextBase TM in a few minutes by importing any relevant text in any relevant language. In addition to quickly building TextBase TMs from legacy “in-house” translation projects, any source of translated text can be easily exploited, including published web content. The potential benefit to translators of being able to easily reference web content is enormous. For example, for a translation project in the field of health care, a translator could quickly import a large quantity of relevant trilingual (English, French, Spanish) content from the World Health Organization web site and begin using it immediately in the translation process. Thousands of other web sites also provide examples of high-quality translations. A noteworthy example is Eur-Lex, an 11-language web site that contains vast amounts of content covering European legislation, treaties, case-law, parliamentary debates, and documents of public interest. A translator working on texts related to food inspection, for example, could simply look up relevant European directives on Eur-Lex, click on the documents in the languages of interest and instantly create a searchable reference TextBase TM.

Translated Expressions of Any Length, In Their Full Context

By providing context, the TextBase TM acts as an extensive "by-example" dictionary, usage and style reference for terms and expressions.

A TextBase TM user can perform a search of the entire TextBase TM for an expression of any length, in any of the languages contained in the TextBase TM. In less than one second, all of the instances of that expression in the entire TextBase TM (in fact, multiple TextBase TMs can be searched simultaneously) are automatically found and retrieved, along with the aligned translation texts. The user can then select and view an instance of the expression and its aligned translation in a split screen view. One half of the view displays all of the text of the document that contains the searched expression, automatically scrolled to the location of the found expression, which is color-highlighted for easy viewing. The other half of the view displays the complete corresponding translation text, scrolled to the aligned text segment, which is also highlighted. At a glance, the translator sees the expression and aligned translation in the contexts of their complete original documents. By providing context, the TextBase TM acts as an extensive “by-example” dictionary, usage and style reference for terms and expressions.

On-The-Fly Alignment: No Up-Front Investment

A point and click environment allows the translator to easily correct the alignment on the fly, thereby improving the TextBase TM continuously with use.

Even with the world's best automatic alignment algorithms, such as the Alignment Agent in MultiTrans 4, which can achieve upwards of 95% alignment accuracy, a complete text in different languages will not always be perfectly aligned. Since the text misalignments are usually off by only a sentence or two, the user can see the problem at a glance in the full-text, split-screen views. A point-and-click environment allows the translator to easily correct the alignment on the fly, thereby continuously improving the TextBase TM with use. Unlike TM databases, the TextBase TM does not depend on perfect alignment; this eliminates the need for time-consuming up-front verification before the system can be used.

An Integrated Translation Support Environment

Once the desired expression and most appropriate translation have been identified from the multiple occurrences, the user can insert the translation into a new translation project. Tight integration of the TextBase TM with a comprehensive translation support environment makes this an easy one-click operation.

Also, the TextBase TM search process can be automated into a batch operation that takes a new translation project, identifies all of the recurring expressions to be translated, then searches and fetches all of those expressions and their translations from the TextBase TM. The resulting suggested translations are then available for very rapid review and insertion into the current project by the translator.

Advanced Terminology Management including: Nominal Terminology, Translation Terminology and Sentence-Level Memory: The Best of All Worlds

If a found expression is a special term that requires formal terminological management, it can be simultaneously added to a multilingual terminology management repository. Terminology management is the domain of skilled terminologists who convert terms into nominal form and manage all of the surrounding information about the term. As with traditional TM systems, an integrated terminology management repository can be used to automatically pre-translate terms in a translation project.

Most recurring expressions (typically 5 words or less), however, are not part of the formal terminology of a subject but are simply sub-sentence units of reusable translation text. These expressions are sometimes referred to as "Translation Terminology" to distinguish them from terminology that is formally managed by terminologists. It is also important to note that this same repository structure allows the seamless incorporation of sentence-level translation memory databases, previously created with traditional TM tools, into the translation process.

Since a translator has validated the translations in the terminology management system and the translation terminology repository, they can both be used for subsequent automatic pre-translation as well as manual search and retrieval.

In addition to building formal and translation terminology repositories on the fly, the expression extraction and alignment tools of the TextBase TM can be used to rapidly build these repositories independent of a translation project.

Productivity and Quality That Improve With Age

By having easy and rapid access to multiple examples of the usage of an expression (of any length) and its previous translations in the context of the full texts to which they belong, translators can quickly solve difficult translation problems with high-quality solutions. Upfront TM database preparation work is eliminated, allowing translators to immediately begin exploiting vast reservoirs of legacy translations. And the interactive nature of the integrated translation support environment allows the TextBase TM, terminology management and translation terminology repositories to continuously improve with usage.

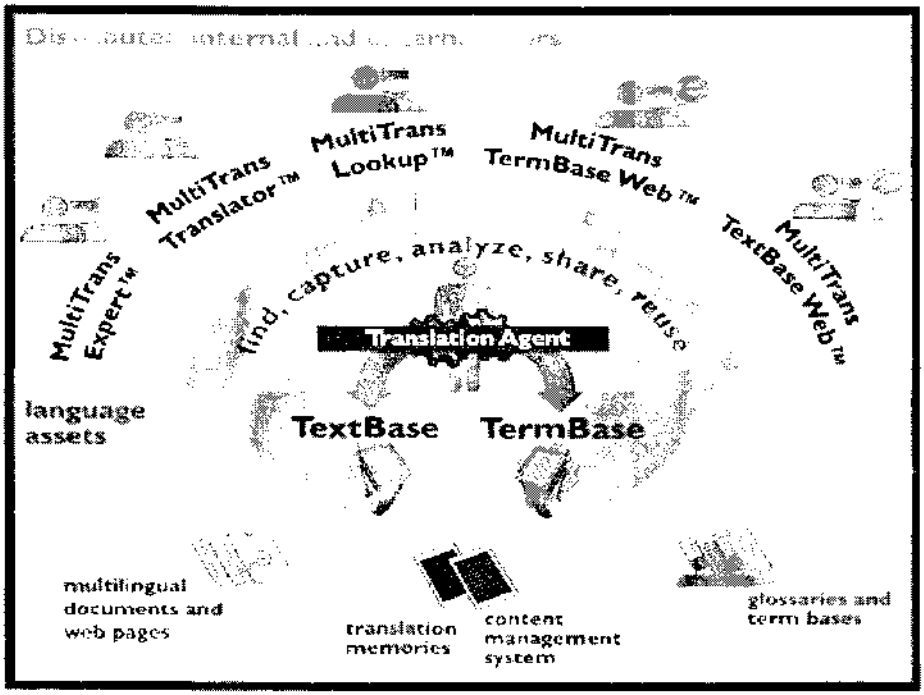
Benefits Across The Information Management Value Chain

As more people interact with the TextBase TM and terminology repositories, they become richer and synergies between translation activity and other information management processes can be realized.

We have focused on the translation process itself; however, the TextBase TM approach offers benefits to many other participants in the information value chain.

The vast repository of expression usage and writing style also speeds and improves the quality of the monolingual new content authoring process by providing authors, reviewers and editors quick access to terminology banks and stylistic, definition and usage guidance from previous content.

Beyond the content-creation process, information consumers can also leverage these valuable resources to solve content-comprehension roadblocks by having easy access to dictionaries and usage examples. For example, if a knowledge worker referring to a company document does not understand a key term, she can quickly look it up in the corpora and/or terminology repositories to see definitions and context-rich examples of meaning and usage. TextBase TMs and terminology repositories become richer as more people interact with them, and synergies between translation activity and other information management processes can occur.



Interoperability with existing TM systems

To fully leverage the TextBase TM approach, interoperability with traditional TM systems and data was crucial. Users can now import their TM's and their terminology databases directly into the TextBase TM and Advanced Terminology Management (TermBase) modules.

Furthermore, at the translation stage, users can process and generate the bilingual unclean files used by the leading traditional TM vendor. So although MultiTrans offers a complete multilingual information management and translation solution, it can easily coexist in environment where more than one tool or approach is being used.

Summary

Hundreds of millions of words of high-quality previous translations exist all around us, including vast resources on public multilingual web sites. Translation productivity and quality could be greatly enhanced if those previous translations could be effectively exploited.

Such has been the promise of Translation Memory (TM) systems; however, traditional TM-based approaches have simply failed to deliver significant benefits. At best, modest gains have been demonstrated in the niche of technical product update documentation - which represents only about 5% of all documents that require translation. For the remaining *95%* of content that must be translated, TM-based systems offer little or no advantage. Their failure stems from their dependence on whole sentence repetition, their loss of translation context and the prohibitive labor-intensity of building the initial TM databases.

Fortunately, a more recent approach to the translation-productivity problem, based on the concept of a searchable full-text multilingual TextBase TM. delivers on the promise. The TextBase TM approach enables the rapid creation of vast pools of

previous translations, provides complete usage and style context for all translations and effectively recycles translations of expressions of any length, not just whole sentences.

The TextBase TM approach provides clearly demonstrable superior productivity for all types of content, including descriptive texts that exhibit no whole sentence repetition. It also helps improve the quality of translations by providing comprehensive “by-example” usage and style references for all participants in the multilingual information-management value chain.

About MultiCorpora

MultiCorpora was the first company to introduce the TextBase TM approach with the release of MultiTrans 1 in 2001. Since then, MultiTrans has evolved tremendously and helped large organizations, corporations, translation companies, and Freelance translators in over 20 countries to reduce the cost of translation while maintaining the quality associated with the art of human translation. Unlike many CAT system providers who are also translation service companies, MultiCorpora is exclusively dedicated to the development and support of the best CAT systems for translators. MultiCorpora is a Canadian organization, with offices in Ottawa, Canada and a European Sales and Services Office in Brussels, Belgium. For more information on MultiCorpora, please visit www.multicorpora.com.

About MultiTrans 4

MultiTrans 4.2 is the latest of the next-generation Computer Aided Translation systems. Using the best of MultiTrans 3 and MultiTrans 4 as a base, along with extensive input from leading corporations, government agencies and translation service providers, MultiTrans 4.2 was designed to enable users to further realize the benefits of the TextBase TM approach. MultiTrans 4.2 provides Advanced Terminology Management capabilities and a robust TextBase TM that recycles 100% of past translations of any segment length and presents them in context for translators. A new TextBase TM Alignment Agent provides industry-leading alignment performance and a new Translation Agent makes MultiTrans 4.2 even easier to use from within the most popular editors such as Microsoft Word. More flexible than ever, MultiTrans 4.2 addresses the needs of all the people involved in the language industry. For more information on MultiTrans 4, please visit www.mu1ticorpora.com.