# Streamlining Multilingual Content Delivery for Large Organizations

## Case *Studies in Enterprise Linguistic Asset Management*

**Bob McCallum**

**MultiCorpora R&D, Inc.**
**Canada**

*bmccallum@multicorpora.com*

## *Introduction*

The language industry is currently facing a knowledge management crisis. The rapidly growing gap between the demand for high quality multilingual content and the lag in the supply of language professionals is driving the requirement for technology that can dramatically boost productivity while improving output quality. In response to this need, MultiTrans provides a linguistic asset management platform that has been proven with large governments, multilateral organizations and multinational corporations to improve multilingual content delivery cycle time and consistency by capturing and exploiting the knowledge of experienced language professionals. The platform includes a centralized platform for sharing multilingual terminology and massive full-text repositories of previously translated document sets among all members of the content delivery supply chain. It also provides an integrated work environment that proactively captures and mines multilingual information that enables language professionals to improve productivity and consistency.

This paper explores the industry trends behind the knowledge management crisis, examines the technology landscape, reviews the MultiTrans platform and illustrates the potential benefits of the proven technology through a set of customer case studies.

## *The Language Industry Knowledge Management Crisis*

### Introduction

The global need for translation productivity and language management technology is poised to explode over the next couple of years. This explosion is being driven by a "perfect storm" of two coinciding trends. On one hand is the growing volume of new, high quality multilingual content that must be delivered by large governments, multilateral organizations and multinational corporations. On the other hand are the lagging supply of experienced language professionals and cost constraints. This rapidly growing gap between the demand and supply of multilingual content delivery capacity will have a profound impact on the abilities of multilingual organizations to achieve their business objectives.

## The Flood of Multilingual Content

Economic globalization, global e-commerce and the expansion of the political mandates of many multilateral public sector organizations are major contributing factors to the growing volume of translation activity globally.

In many industries, including pharmaceuticals, financial services, legal services, automotive and many other sectors, large companies are merging to create truly global corporations. These organizations increasingly serve a multilingual constituency of employees, investors, partners and customers. With the removal of many trade barriers over recent years, the number of small and medium sized businesses that serve global markets is also rising rapidly.

The Internet, including international e-Commerce, is a specific aspect of globalization that is fueling global translation activity as the number of web users with diverse native languages rises. According to the prominent industry analyst Aberdeen Group:

- Non-English speaking web users are outpacing English users, increasing from 52% in 2000 to 68% in 2003.

- By end of 2003, 66% of e-commerce transactions will originate outside of the US.

- Web surfers are 3 times more likely to purchase on a web site in their preferred language.

On the public sector front, many governments and multilateral organizations are rapidly expanding their mandates to serve stakeholders in their native languages. The European Union is expanding from 15 to 25 member countries in 2004. Many United Nations agencies are also expanding the scope of their mandates related to multilingual content delivery.

In summary, the corporate and public sector organizations of the world are creating content that must be served in multiple languages in greater volumes than ever before and the growth in that volume is accelerating.

## The Human Capital Drain

At the same time that demand for multilingual content is growing rapidly, the industry is at the beginning of a major global stagnation in the number of language professionals. The demographic profile of the average translator is that of an aging baby boomer within 5-10 years of retirement. The number of young translators entering the profession has been insufficient to compensate for this attrition and satisfy the growing demand.

A July 2002 Industry Canada study, entitled "The Language Industries in Canada" provides the following quote to illustrate the growing gap between the supply and demand of translation services. *"There will not be enough new professionals to bridge the widening gap between supply and demand. According to several indicators, the shortage of new people entering the field, especially in translation, appears to be global."*

Also, it is not a simple numbers game of replacing retiring language professionals. Writing and translation are knowledge-intensive activities. Translators 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, linguistic skill and deep subject matter expertise. Those capabilities develop as the accumulated knowledge in the heads of experienced translators. When they retire or leave an organization, that valuable knowledge leaves with them. Without an efficient way to capture and transfer that valuable knowledge – an important enterprise linguistic asset – new, more junior professionals must start from scratch, slowly learning and accumulating that same knowledge for themselves. With the coming wave of language professional retirement, many organizations will experience a massive loss of knowledge during the next few years. In some organizations this is of crisis proportions: take the actual case of one multilateral organization department of approximately 20 language professionals that will lose over 80% of the existing team to retirement within 7 years.

## Supply Chain Complexity

Complexity in the multilingual content delivery supply chain compounds the productivity gap problem. For large organizations, there may be thousands of distributed source-language authors, in-house referencing and translation teams distributed across multiple locations or telecommuting from home offices, documentation editors and revisers, and external service providers - both agencies and freelancers. There are also the tens of thousands of knowledge workers inside and outside of the organization who are the ultimate consumers of the multilingual content.

This complexity creates many opportunities for redundant activity, inconsistent and incorrect use of language, and excessive revision time and cost.

## The Business Impact

The impact of the productivity gap and supply chain complexity is profound. Multilingual content delivery is becoming a bottleneck to globalization, slowing time to market. Also, when resources become stretched and large projects get split across teams of language professionals to improve turn-around, the consistency and quality of the communications suffer. Excessive editing and rework also contribute to excessively high communications costs.

The impact on multinational corporations is measured in millions of dollars of revenue for every day of delay in the delivery of new products to international markets. In the pharmaceutical industry, for example, the rollout of new drugs to hundreds of countries involves a very large volume of critical documentation for the regulatory approval process and subsequent marketing launch. By shrinking the time to accurately translate that content into the local language of each country, overall time-to-market is reduced dramatically in an industry where the opportunity cost of delays is measured in millions of dollars of drug revenues per day in a single foreign market.

For governments and multilateral organizations, the failure to deliver high quality content in a timely and cost-effective manner to multilingual constituents leads to high-level political fall-out.

The situation is critical and the opportunity is ripe for technology to bridge the productivity gap, streamline collaboration in the face of supply chain complexity and solve the knowledge retention problem.

## The *Technology Promise*

## The Language Technology Landscape

In addition to content management, project management and workflow management systems, there are three major types of core translation productivity and automation technologies: 1) Machine Translation (MT); 2) Translation Memory (TM); and 3) the full-text multilingual search agent approach used by MultiTrans, the MultiCorpora product. The diagram below plots the area of coverage of each technology, considering the range of content types each technology addresses and the related quality of translation supported.



**Translation Technology Landscape**
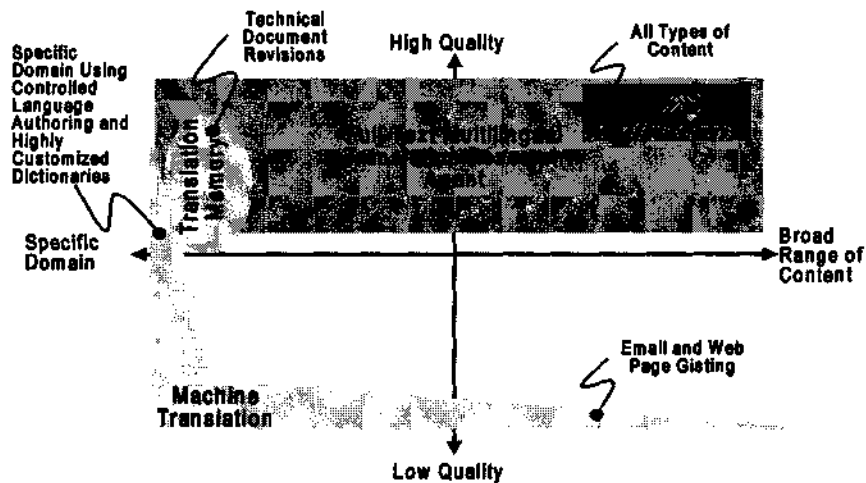
## Machine Translation: Low Quality "Gisting" and Specific, Controlled Authoring Applications

Machine Translation involves the use of complex computational linguistic algorithms to automatically (without human translator involvement) translate text from one language to another. Linguistic rules aim to dissect the source text, identify its grammatical structure and understand its vocabulary, then reconstruct the same information from the vocabulary and grammar rules of the target language. It is an immensely difficult activity that relies on huge volumes of rules and dictionaries - which are specific to each language pair being handled.

Machine translation technology has been an area of academic and commercial pursuit for approximately 50 years. In the past, proponents envisioned machine translation as the eventual replacement for human translators for most types of translation activity. It is now generally accepted that machine translation is useful in only two specific applications. The most common application is for producing a very rough translation that is highly incorrect but provides the "gist" of the content under consideration. In

situations where content is not available in the desired language, and where the investment to accurately translate using human translators would not be made, it provides a valuable capability. Individuals and corporations broadly use machine translation (including free Internet based services such as found on altavista.com) to "gist translate" email and web pages. Here, the quality is poor but better than no translation at all. Because the content being gisted is not a likely candidate for human translation, machine translation is entirely complementary to human translation activity.

The second application of machine translation is in very specific domains where a company has made a considerable investment in 1) the authoring process to control the use of grammatical structure and vocabulary to improve machine understandability and 2) the development of the linguistic rules and custom dictionaries to make the machine translation system reliable. Even in these cases, human post-editing follows machine translation. Examples of the use of machine translation exist in the largest aerospace and automotive companies for technical documentation translations for specific products between specific language pairs.

## Translation Memory: Addressing Only 7.5% of the Translation Universe

As an alternative to the linguistic approach of Machine Translation, conventional Translation Memory (TM) systems are based on the concept of a database of previously translated sentences that can be searched to identify and reuse the translated text for sentences that recur in new translation projects. The concept is powerful: by reusing previously translated segments rather than repeating the same translation from scratch every time it occurs, productivity and consistency are increased.

However, because TM systems rely on a database of isolated whole sentences, their usefulness is limited to translation projects where many whole sentences in a new source document have been previously translated in other documents. This occurs for technical documents that are frequently revised – for example, the user manual for a Pentium 3 computer likely reuses many sentences from the user manual for the earlier Pentium 2 model.

Unfortunately, the fraction of global translation activity that involves a sufficient level of recurring whole sentences where TM can provide benefits is estimated to be only about 7.5%, as illustrated below.
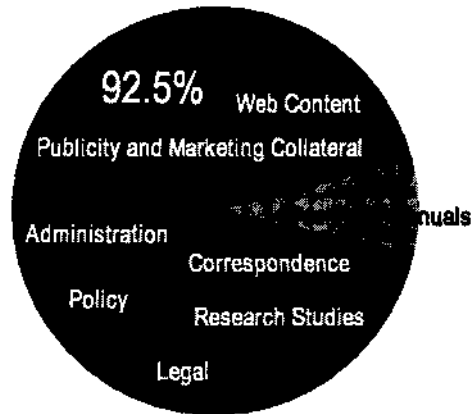
For the remaining 92.5% of translation activity (reports, correspondence, marketing literature, etc.), the content involved exhibits very little sentence repetition but does possess a high level of repetition of sub-sentence expressions and phrases. Due to these constraints, the translation process is not adequately supported by TM-based approaches. It is not surprising, then, that TM tools have not surpassed a market penetration rate of about 10% despite being actively promoted for many years.

Another drawback of TM systems is the requirement for a database of perfectly aligned sentences before the system can be used. Perfect alignment requires manual verification and correction – even when the most sophisticated auto-alignment tools are used. This manual pre-alignment process is very tedious and time-consuming – it typically takes weeks of effort to develop modest size TMs. Also, research shows that less than 30% of a TM database is ever referenced; however, 100% must be aligned since it is impossible to predict which sentences will be recycled. Many organizations simply refuse to invest the time needed to build new TM databases and valuable knowledge remains locked in legacy documents.

Finally, authors, translators and editors all need contextual information to be able to make informed terminological decisions. Since translation memory tools work with a database of isolated aligned sentences, retrieving contextual information – a looking at the segment within its complete sentence, paragraph, or document – becomes impossible.

## MultiTrans: A Linguistic Asset Management Platform for Large Multilingual Organizations

In order to bridge the productivity gap and streamline the complex multilingual content delivery supply chain, MultiTrans was developed on a foundation of five key elements:

1. A massive, searchable repository of the aligned full-text of previously translated document sets

2. An industrial-strength terminology management system

3. An integrated translation support environment that puts relevant references at the fingertips of a translator in familiar editing environments and allows the knowledge repositories to learn and grow with usage

4. An intelligent software assistant to the translator that proactively automates and aggregates searches and comparisons of new translation projects with all relevant sources of translation examples, approved terminology and legacy translation memory files

5. A centralized platform for the sharing of linguistic assets along the entire multilingual content supply chain, from source language authors to multiple target language content consumers
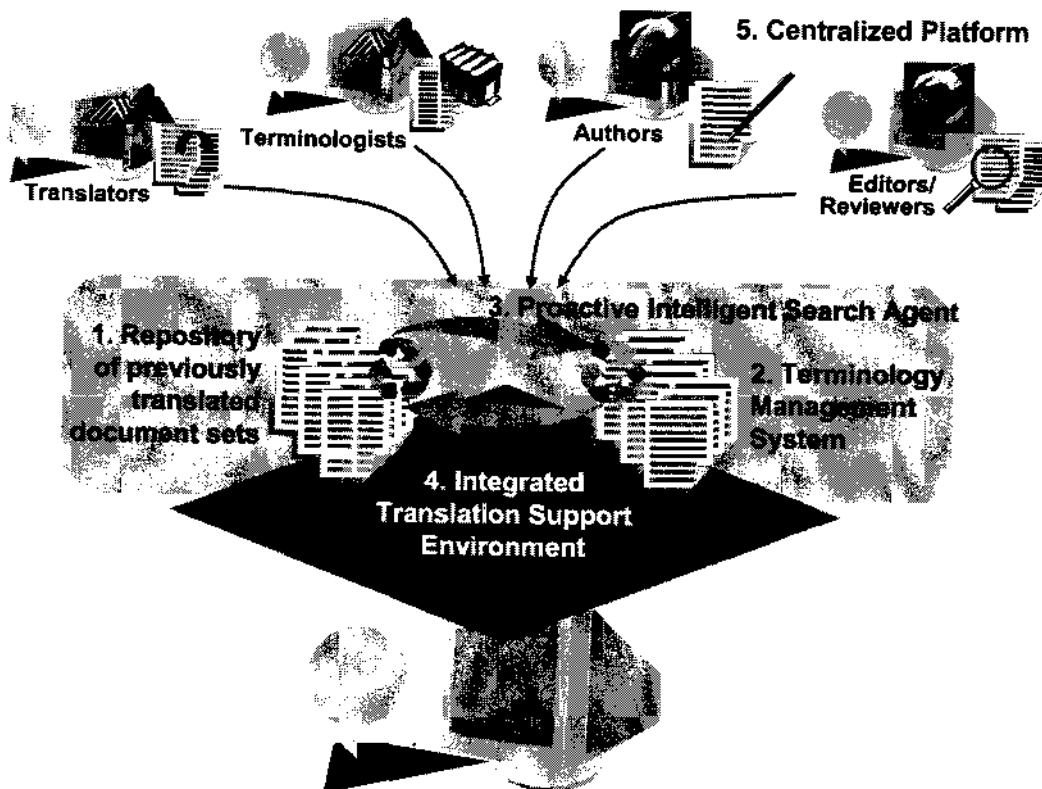
## A Full-Text Multilingual Repository of Previous Translations

MultiTrans enables users to easily build and maintain indexed repositories of cross-referenced multilingual (or monolingual) content from legacy documents. The system automatically aligns legacy translated document pairs based on user-defined file nomenclature rules. It then automatically extracts, indexes and aligns the full text. The full-text indexing enables text strings of any length (words, sub-sentence expression, whole sentences or paragraphs) to be subsequently searched and retrieved. Sophisticated statistical alignment algorithms establish links between the equivalent sentences in multiple language versions of the same content. When an expression of interest is found in one language within the repository, the alignment link allows the corresponding previous translation to be also identified and retrieved.

When an expression and its aligned other-language version are retrieved, they are presented to the user in side-by-side scrollable windows showing each expression highlighted within their surrounding full text, providing the user with valuable usage and style context. Since automatic alignment algorithms cannot yield perfect results all of the time (for example when a sentence in one language is split into two during translation), the full context views of MultiTrans allow a user to spot a misalignment at a glance (on the occasions when there is a misalignment, it is only off by one sentence in over 95% of cases) and correct it on the fly. Because of this context and ability to correct the occasional misalignment as they work, translators can begin using a repository within minutes of its automatic creation and the repository actually improves over time with usage.

With MultiTrans, a very large searchable repository of previous translations can be built very rapidly – at a rate of approximately 50,000 words per minute on a low-end computer. A corpus of hundreds of millions of words can be built in a matter of hours and be ready for immediate use by language professionals.

## Terminology Management

Large organizations face the challenge of maintaining a consistent corporate "language" in all of their communications, be they mono or multilingual. When multiple functions or geographic regions within an organization work together on efforts such as product development or marketing, terminology inconsistencies grow rapidly and become a source of frequent rework and costly delays. The consistent use of proper terms (and the avoidance of use of undesirable terms) is also critical for communications clarity and accuracy and the reinforcement of an organization's global brand.

Since a MultiTrans full-text repository provides multiple examples of previous full text translations, it serves as a valuable reference that reinforces a standard corporate language. MultiTrans also provides "terminology management" capabilities to complement the full-text repository with information on which special terms and translations have been specifically reviewed and approved for usage. Terminology management, as a discipline, involves the identification of candidate terms from legacy and new documents, research into the usage, meaning and translations of terms, approval of the use of the term and its translations, and the sharing and dissemination of approved terminology across all potential users.

In MultiTrans, managed terms can include nominal form terminology (as used by terminologists), any pre-approved translation of a word, expression, sentence or paragraph (sometimes called translation terminology), or translated sentences from Translation Memory files. Terminology and equivalencies from many external sources can be easily incorporated into the MultiTrans platform, including existing corporate dictionaries, industry glossaries and classification references, and third-party terminology bases such as Termium.

MultiTrans supports the terminology management process with a scalable platform for storing and tracking comprehensive terminology management information and a suite of integrated capabilities to:

- Enable users to easily capture new terminology during the authoring or translation process;
- Automatically extract terminology candidates and their corresponding translations from legacy documents;
- Provide supporting tools and resources to support terminology research activities;
- Manage the approval status and lifecycle of terminology; and
- Make approved term translations available to translators for automated or manual inclusion in new translation projects via the integrated MultiTrans workbench.

## An Integrated Translation Workbench

While MultiTrans provides dedicated user interfaces to directly manage and search the multilingual full-text repository or terminology repository, it also provides a tightly integrated environment within Microsoft WORD or other popular editing environments (PowerPoint, WordPerfect, an HTML editor and others) where all of the language assets and search functions are available to a translator from within the documents that they are working on.

Starting from the document to be translated (in its source language, say English), the translator interacts with MultiTrans functions to rapidly find examples of previous translations (in the target language, say French) of matching expressions, sentences and terms from the relevant multilingual reference corpora and terminology repositories. Previously translated segments that are selected for inclusion in the project are inserted into the document by overwriting the English text with the French translation. In this way, the document evolves from all English at the beginning of the project to a hybrid document where some of the text remains in English and some has been translated into French based on examples of previous translations found by MultiTrans. The translator then completes the project by manually translating the remaining English portion.
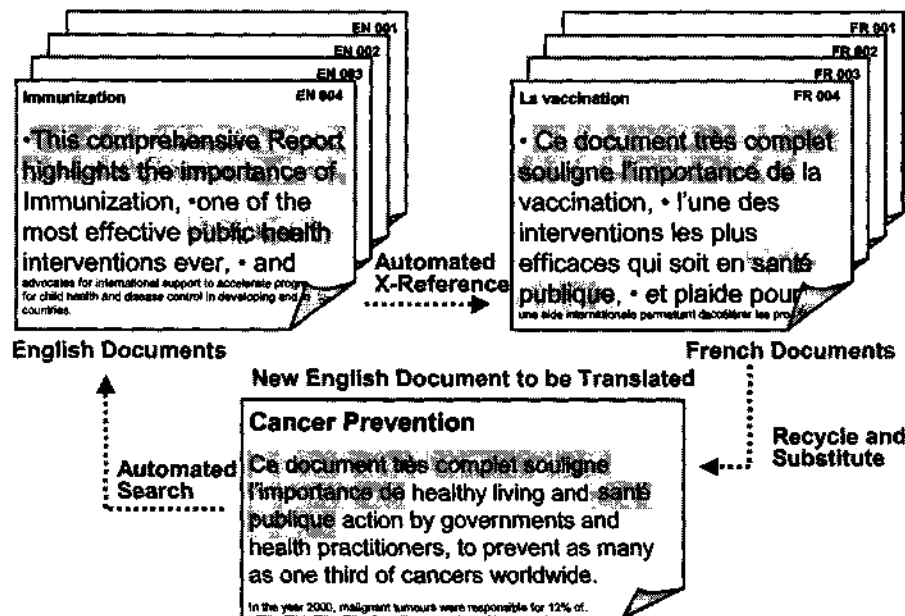
Within the MultiTrans tightly integrated working environment, the multilingual full-text repository and terminology banks learn and grow continuously with usage. As full-text translation examples are retrieved and reused, minor misalignments in the repository are automatically corrected on the fly – the more the repository is accessed, the better the alignments become. As expressions are retrieved from the full-text repository, they can be automatically inserted into the terminology base, allowing it too to learn and grow with translator usage. Also, as new expressions are manually translated during a project, the capture of that term and its addition to the terminology base is a single mouse click action.

# Integrated Translation Support and Language Management
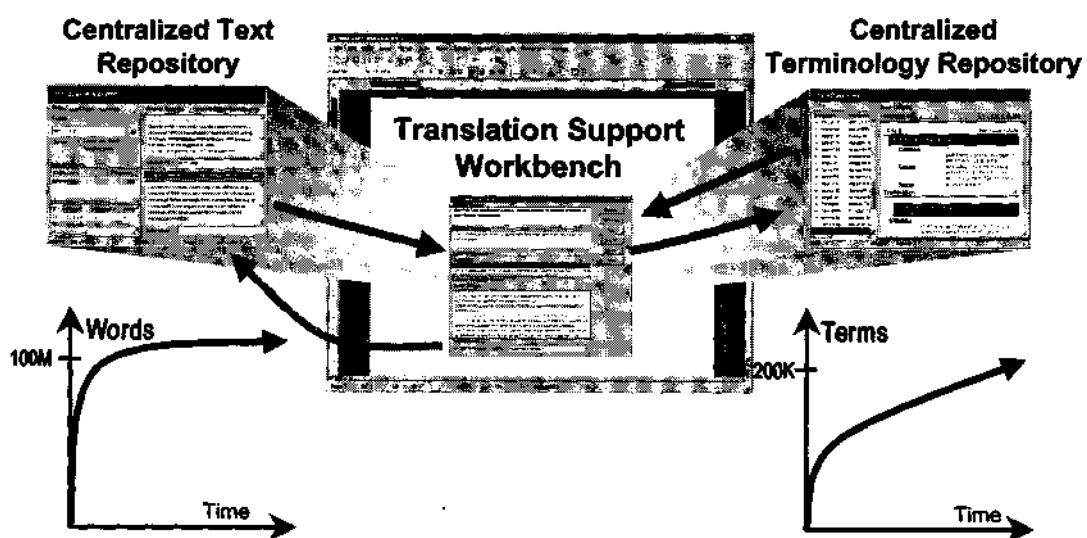


As new content is created and translated, it is added to the full text repository, expanding the example base for future users. The text volume curve above shows how the repository grows very quickly as a result of the initial corpus building exercise to

incorporate legacy translations. Then, over time it grows slowly and steadily as new translation work is completed.

The term volume curve shows that an initial terminology repository can be created very quickly from legacy term bases and automated terminology generation – MultiCorpora also offers an optional terminology generation product. Then, as translation activity occurs, new terms are identified and captured, allowing the terminology repository to continue to grow rapidly.

From the workbench, MultiTrans provides a number of types of search capabilities to mine full-text repositories and terminology repositories for examples of translations. An automatic pre-translation mode compares an entire new project in one batch operation to approved terminology repositories and automatically retrieves and inserts the corresponding translations. The translator can also manually select any expression in a new project and execute a search of the full-text and terminology repositories. More commonly, translators use a single automated and optimized multilingual search and comparison process that combines all search methods into one user action.

## Automated and Aggregated Multilingual Search and Comparison

Deploying a sophisticated search and comparison algorithm, MultiTrans automatically compares an entire new source document to all of the open full-text content and terminology repositories and proposes the set of matches that maximizes the suggested reuse of previous translation work. It does this by considering full sentence matches (exact and fuzzy), terminology matches, and thousands of expression matches covering all possible combinations of words. A sophisticated algorithm prioritizes the results and presents them to the translator according to what will provide maximum reuse of previous translations while following rules about the relative quality of the source – certain document repositories can be rated above others, exact matches with approved terminology have priority over full-text repository matches, etc. This whole search process is one step and executes in less than a second – even when referencing sources that total millions of words. As always, the full original usage and style context of the matched expressions and their equivalent translations are provided. Reviewing and inserting the suggested translations is then a simple point-and-click operation for the translator who remains in control of the entire process.

Interactive Translation Research Agent
Aggregation, Automation, Context

The previous figure illustrates how this aggregated, automated search capability mines many sources of potentially valuable multilingual references, including:

- Previously translated original documents in WORD, HTML, PDF, or other formats;
- Multilingual web sites, such as the WHO trilingual site with thousands of words of high quality translations in the health field;
- Other sources of previously translated documents;
- Existing Translation Memory databases. In fact, because MultiTrans fully indexes these files too, users get all of the benefits of TM exact and fuzzy full sentence matches plus the ability to mine deeper by retrieving sub-sentence strings. MultiTrans also creates industry-standard format TMs as an output of the translation process. Combining this with the other search capabilities makes MultiTrans a superset of TM systems.
- Terminology databases, project-specific lexicons, industry classification references and third-party glossaries.

## A Platform for Sharing Linguistic Assets

The multilingual full-text and terminology repositories are linguistic assets that can be shared in real-time across the entire enterprise and supply chain via a centralized platform that enables workstation or web browser access from anywhere, anytime.

A centralized and web-based environment enables, over a network or the Internet, several users to search, share, view and update the same full-text or terminology repository in real-time – the moment a user adds a term or corrects an alignment, that

new language management asset can instantly be accessed and used by other language professionals - both those with a MultiTrans workstation and those that are accessing in a read-only mode via a standard web browser.

## Real-Time Enterprise Collaboration
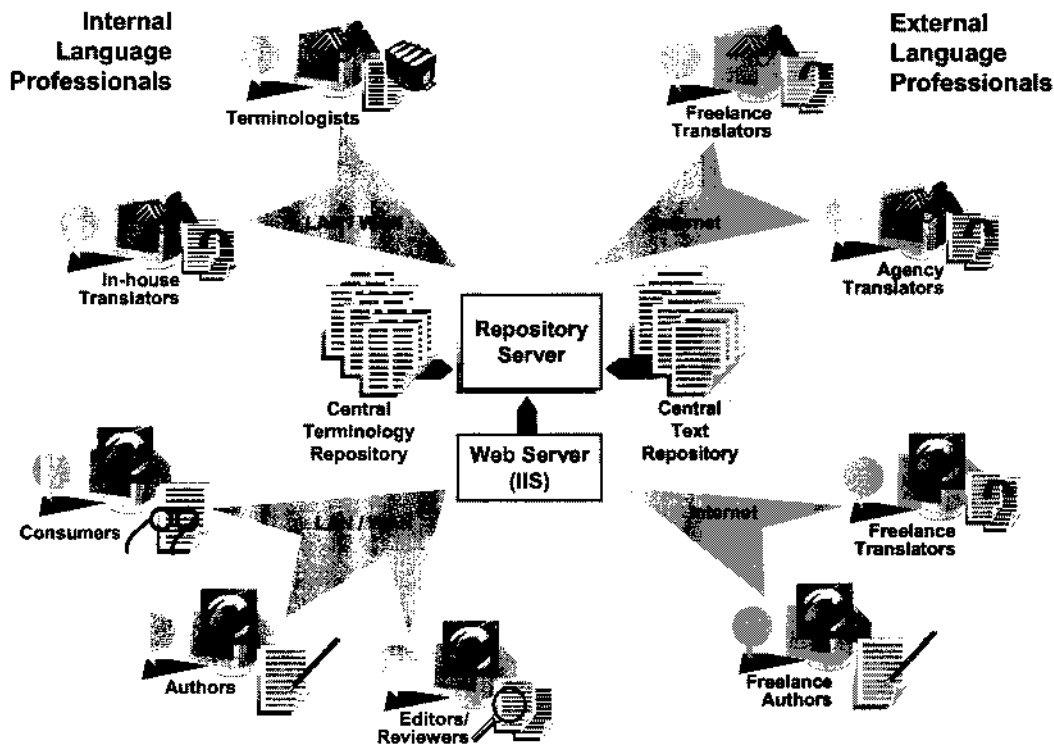


Authors can reference the central language resources for monolingual assistance with term definition, style and usage and full-text examples of previous usage.

**Teams of Translators** distributed across in-house, agency and freelance organizations can share central linguistic assets.

**Terminologists** can leverage the resources for new terminology generation and research support during translation projects.

**Editors and Reviewers** can clarify terms, usage and see examples of previous translations.

**Content Consumers** who need to read and understand documentation can obtain clarification on terms and language by easily accessing definitions and examples.

This flexible platform easily integrates with a variety of content management and workflow system environments. It also supports and streamlines the wide variety of workflow patterns found in large complex organizations, including:

- The use of centralized document analysis and referencing prior to dispatch to translation teams
- Project-based new terminology extraction and standardization prior to dispatch to translation teams

- Staff located across multiple geographic locations working on corporate local or wide area networks or telecommuting from home offices using dial-up Internet connections
- External service providers, either with MultiTrans workstations or standard web browsers

The MultiTrans centralized platform links all participants in the language management value chain to one central repository of information, eliminating duplication and rework and shrinking content delivery cycle time.

## *Case Studies*

The MultiTrans solution, based on the five foundation elements described above has been proven over several years with a wide variety of organizations, including leading governments, multilateral organizations and multinational corporations. The following organization case studies, based on actual MultiTrans customers, illustrate the challenges faced by large multilingual organizations and how the deployment of the MultiTrans platform enables productivity and quality breakthroughs across the multilingual content delivery supply chain.

## Case Study #1: A Large International Utility Company

### The Challenge

A large and growing utility company with annual revenues of $13 billion from operations in 50 countries and assets of almost $60 billion embarked on a major capital investment expansion in 2003 to increase capacity, reinforce system reliability and improve customer satisfaction. In the face of major changes in the regulatory environment, a central strategy of the company is to increase profitability while maintaining excellent customer service through information technology innovation.

With global operations and the initiation of major capital projects, the requirement for the company to deliver multilingual content is growing rapidly. A large team of over 60 translation professionals handles a wide variety of document types and sizes, including annual reports, project information, customer communications, marketing and public relations, human resources and administration, and technical information. The translation services group serves approximately 25 distinct internal customer groups, each with thousands of previously translated documents and extensive terminology files related to each area. The increase in translation activity as a result of company growth and major capital projects combined with the corporate initiative to increasing profits and customer satisfaction have put considerable pressure on the translation unit to maintain quality while reducing costs and increasing throughput.

These challenges are heightened by the fact that translation staff members telecommute from home offices on a regular basis and the group makes extensive use of external translation services providers. With an ageing workforce, the translation unit is also expecting significant staff turnover during the next few years due to retirement attrition.

### The Solution

The translation services group recognized that the promise of dramatic increases in productivity and quality lie in the use of state of the art information technologies to capture and leverage their linguistic assets – which were locked up in thousands of previously translated documents sets. With global operations, telecommuters and distributed external translation service providers, the company wanted a controlled, centralized way to share linguistic assets and standardize terminology usage by all the people involved in creating and delivering multilingual content.

After a comprehensive review of available translation memory and document search technologies by an independent consultant, the group selected the MultiTrans platform. A key factor in the selection was the MultiTrans ability to provide a reliable and secure centralized platform where terminology and full-text multilingual

repositories could be hosted remotely by MultiCorpora and accessed by MultiTrans workstations or standard web browsers via the Internet. Other important considerations were the user-friendliness of the integrated translation workbench working within Microsoft WORD, the ability of MultiTrans to work equally effectively in any language direction (e.g. French to English as well as English to French), and the excellent customer service experience that established the faith in MultiCorpora as a strategic technology partner that can deliver the support and a continuous stream of language technology innovations that will enable the group to continuously improve performance in the future.

The centralized deployment of MultiTrans is illustrated below. Approximately 25 full-text repositories and 25 terminology banks are hosted off-site by MultiCorpora. The company's distributed language professionals and external translation service providers are able to access and share the centralized linguistic assets either through a standard web browser or seamlessly through the MultiTrans workbench, primarily through Microsoft WORD and PowerPoint.



**Centralized MultiTrans Deployment**
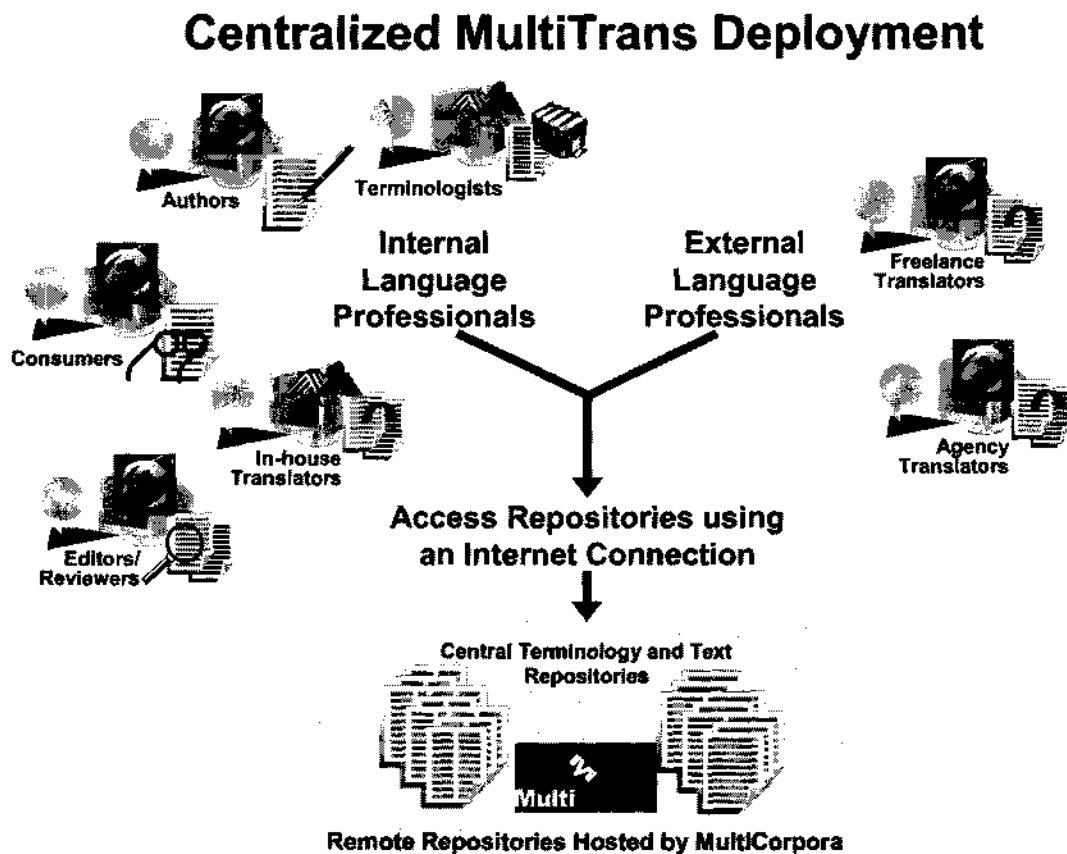
**The Impact**

An analysis of the types of multilingual documents that the company produces is summarized below. Typical of a large multinational with a broad mix of documents sizes and types, whole sentence repetition is quite low at 13%; however, sub-sentence expressions that repeat and can be reused from previous translation projects account for approximately 42% of new documents. The resulting translator productivity gain is approximately 35%.

In addition to the simple translator productivity gain are the quality gains due to terminology standardization and the related cycle time and cost savings from more consistent source language authoring and reduced translation revision effort.

| Section I: Content Characteristics | |
| --- | --- |
| **Content Characteristics** | **Values** |
| Exact match rate of whole segments (whole sentences, titles, table entries, etc,.) in new document compared to corpora of previous translations | 13% |
| Exact match rate of terms in new document compared to terminology repositories | 12% |
| Fuzzy match rate (85-99% found) of whole segments | 15% |
| Exact match rate of short expressions (at least 2 consecutive words and typically less than 5 words within a segments where the overall fuzzy match rate is less than 85%) in new document compared to corpora of previous translations | 26% |
| New content remaining that requires manual translation | 34% |
| Total | 100% |

| Section II: Automated Process Efficiencies | | |
| --- | --- | --- |
| | **Efficiency (% of Manual Effort to Process Translations)** | |
| **Automated Process Efficiencies** | **Values** | **% of Content** |
| Exact match whole segments | 30% | 13% |
| Translation Memory Fuzzy Match (85%) | 50% | 12% |
| Exact match terminology | 40% | 15% |
| Exact match expressions from corpus | 60% | 26% |
| New content | 100% | 34% |

| Section III: Productivity Gain | |
| --- | --- |
| **Productivity Gain with MultiTrans** | **35%** |

## Case Study #2: A Large Government Department

**The Challenge**

In response to a critical knowledge attrition problem and a strategic cost reduction imperative, a large government department is deploying the MultiTrans platform to retain and leverage the knowledge of senior language professionals while significantly reducing the cycle time and costs associated with the delivery of multilingual content.

With an annual operating budget of over $3 billion, a rapidly growing employ base of over 50,000 employees at 750 sites who respond to over 28 million public enquiries per year, the department is facing a knowledge management crisis with the eligible retirement of over 43% of its executives and 16% of all of its employees within the next 7 years. Also, the department's strategic plan calls for annual operating cost savings of $50 million as a result of increased productivity and reduced administrative costs through greater use of electronic processes and automation.

A large group of translators, distributed across multiple teams within the department, deliver translation services to the growing department. The department also outsources a considerable amount of overflow work to external translation service providers. Like most government organizations, the department generates a wide variety of multilingual documentation and correspondence that is non-repetitive in nature. Detailed analysis revealed that repeating whole and fuzzy sentence matches account for less than 6% of the text volume for over 75% of the department's documents.

With the growing volume of non-repetitive documents, employee attrition and cost-reduction pressure, the translation services units of the department were faced with the challenge of significantly increasing throughput while maintaining high quality.

**The Solution**

The department evaluated and rejected the leading Translation Memory (TM) systems because of the low incidence of whole sentence repetition in their documentation. Also, a single unit within the department had over 6,000 pairs of previously translated documents containing over 12 million words of text that it wished to make available for automated referencing. They concluded that to create comprehensive translation memory files containing those documents would take years of effort, which was simply prohibitive.

After careful evaluation, MultiTrans was selected because of its ability to rapidly build massive full-text repositories of previously translated documents pairs that enable the recycling of text strings of all lengths. The linguistic services officers that reviewed MultiTrans also praised its user-friendliness.

**The Impact**

Despite a very low whole sentence repetition rate (less than 6%) in the majority of their documents, MultiTrans enables the unit to mine sub-sentence expressions that have been previously translated that cover over 16% of the text volume in their typical documents. The detailed breakdown of the documents characteristics of this department and the productivity gains enabled with MultiTrans are summarized in the

following table. Overall, the document characteristics of this typical government department correspond to a recycling productivity gain of over 20%.

| Section I: Content Characteristics | |
|---|---|
| **Type 1: Administrative Correspondence, Marketing Materials, Corporate Web Sites, and Other Content Without Many Repetitive Whole Sentences** | |
| **Content Characteristic** | **Values** |
| Exact match rate of whole segments (whole sentences, titles, table entries, etc,.) in new document compared to corpora of previous translations | 4% |
| Exact match rate of terms in new document compared to terminology repositories | 2% |
| Fuzzy match rate (85-99% found) of whole segments | 2% |
| Exact match rate of short expressions (at least 2 consecutive words and typically less than 5 words within a segments where the overall fuzzy match rate is less than 85%) in new document compared to corpora of previous translations | 19% |
| New content remaining that requires manual translation | 73% |
| Total | 100% |

| Type 2: Technical Manuals, Catalogs, and Other Content With Repetitive Whole Sentences | |
|---|---|
| **Content Characteristics** | **Values** |
| Exact match rate of whole segments | 40% |
| Exact match rate of terms | 5% |
| Fuzzy match rate (85-99%) of whole segments | 20% |
| Exact match rate of short expression | 20% |
| New content remaining that requires manual translation | 15% |
| Total | 100% |

| Breakdown of Workload by Type of Content | % of Total Workload[1] |
|---|---|
| Type 1: Non-Repetitive Whole Segment Content | 75% |
| Type 2: Repetitive Whole Segment Content | 25% |
| Total | 100% |

| Section ii: Automated Process Efficiencies | | |
|---|---|---|
| | **Efficiency (% of Manual Effort to Process Translations)** | |
| **Automated Process Efficiencies** | **Values** | **% of Content** |
| Exact match whole segments | 15% | 13% |
| Translation Memory Fuzzy Match (85%) | 40% | 3% |
| Exact match terminology | 40% | 7% |
| Exact match expressions from corpus | 70% | 19% |
| New content | 100% | 59% |

| Section III: Productivity Gain | |
|---|---|
| **Productivity Gain with MultiTrans** | **22%** |

## *Summary*

The translation industry is currently being affected by two opposing trends: the rapid growth in demand for high quality multilingual content and the lag in the capacity of language professionals to keep pace. The impact of the productivity gap is profound.

Multilingual content delivery is becoming a bottleneck to globalization, slowing time to market. Also, as resources become stretched and large projects split across teams of professionals to improve turn-around, consistency and communications quality suffer. Excessive editing and rework also contribute to excessively high communications costs. This rapidly growing gap between demand and supply is driving the requirement: for technology that can dramatically boost productivity while improving quality.

A successor to Translation Memory (TM), the MultiTrans linguistic asset management platform based on a full-text multilingual content repository has been proven for all types of content across multiple domains, such as the pharmaceutical industry, across all levels of government, multilateral organizations, and other business sectors worldwide.

Based on the content delivery profile of leading multilingual organizations, the return on investment from this technology is very large and readily measured. The translation productivity gains alone are 20 to 80% and the associated payback is less than 180 days. Implementation, integration and training are minimal and non-disruptive to the customer's IT organization. In addition to cost savings, MultiTrans enables increases translation throughput, reduces turn-around time and improves communications consistency.