



Terminology Management in the Localization Industry

Results of the LISA Terminology Survey

Kara Warburton, LISA Terminology Advisor
October 16, 2001

Terminology Management in the Localization Industry.....	1
Results of the LISA Terminology Survey.....	1
Introduction.....	3
Methodology.....	3
Presentation of the Results.....	3
Definition of Terms.....	4
Findings – Survey 1.....	5
Observations.....	6
Findings – Survey 2.....	7
Observations.....	10
Findings – Survey 3.....	11
Observations.....	23
Overall Conclusions and Recommendations.....	26
Non-Communications Industries.....	26
Translation and Localization Services and Organizations That Have a Termbase.....	26
Developers of Translation and Localization Tools.....	28
Closing Remarks.....	28
References.....	30

Introduction

The Localization Industry Standards Association (LISA) provides best practice advice, business guidelines, and standards information for translation and localization workflow, tools, and enterprise globalization. In support of this mandate in the area of terminology management, LISA established a Terminology Advisory position which was held in 2001-2002 by Kara Warburton, a terminologist at IBM.

To fulfill her mandate, Ms Warburton conducted a survey on behalf of LISA to research the methods, tools, and practices for managing terminology in the localization industry. Over 800 representatives from a range of industry sectors worldwide that are involved in globalization were invited to participate. This document presents the findings of the survey.

This document is being distributed to LISA members and to the respondents of the terminology survey.

Methodology

The localization industry is represented by a diverse range of organizations, such as translation service providers and their clients, software developers, global industries in sectors such as telecommunications and medicine, and even governments, trade unions and academic institutions. In order to ask questions that were tailored to different respondent types, three different surveys were offered. Respondents were asked to complete the survey that best met their profile:

- Survey 1 – for organizations that do not perform any terminology work (7 questions)
- Survey 2 – for organizations that perform terminology work but do not maintain a termbase (28 questions)
- Survey 3 – for organizations that perform terminology work and also maintain a termbase (43 questions)

Presentation of the Results

This report summarizes the findings of each survey. After the statistics and comments, a section titled “Observations” proposes some additional conclusions that may be drawn from these findings. These conclusions are sometimes based on patterns, extrapolations, and interpretations and are not always demonstrable by the statistics.

For simplicity, percentages have been rounded to the nearest multiple of five unless it was felt that this would unbalance the results. This can result in total percentages not equaling 100.

Definition of Terms

To avoid any misunderstanding about the terms used in the survey, the following definitions were provided:

concept

A unit of thought constituted through abstraction on the basis of properties common to a set of objects. A concept can be designated by a term and described by a definition.

context

Text or part of a text in which a term occurs.

corpus

A body of text.

localization

The process of modifying products or services to account for differences in distinct markets.

terminology work

Any activity whose goal is to ensure that the use of terminology is appropriate and correct. Terminology work includes the following:

- Collecting terms, for example, for glossaries, dictionaries, word lists, or terminology databases
- Comparing terms, for example, to identify synonyms, related terms, or variants
- Describing terms, for example, in preparing definitions, explanations, or usage information
- Cataloging or organizing terms, for example, assigning subject fields or other administrative categories
- Archiving terms in a terminology database
- Distributing terms to employees, customers, or other stakeholders
- Prescribing terms through a standardization or sanctioning process
- Monitoring term usage, for example, through quality-control activities (editing, testing, screening, etc.)
- Creating new terms (neologisms) for new concepts

language processing tool

An automated tool used to perform terminology work.

source language

The language that is used in the original version of a product's assets (documentation, user interface, packaging materials, etc.) before translation to other languages.

target language

The language into which a product's assets are translated.

term

Designation of a concept in a special domain (science, technology, economics, etc.) by a linguistic expression. A term may consist of one or more words.

term entry

The set of information describing a term, such as the term itself, definition, context, part of speech, translation, related terms, and so forth, and usually stored in a terminology repository such as a termbase.

termbase

A collection of terms and information about terms, such as translations, definitions, grammatical information, usage information, and contexts. In contrast to a simple word list, a termbase contains a range of different data about terms and is frequently maintained using a software program or database management system.

Findings – Survey 1

The main goal of Survey 1 was to find out which types of organizations do not perform terminology work, why they don't perform terminology work, and what benefits they might realize if they did perform terminology work. Twenty-three responses were received. Thus, 30 percent of the respondents of the terminology surveys do not conduct any terminology management whatsoever.

The majority of the respondents to this survey (70 percent) are in the software development industry, while less than 10 percent are in the translation industry. This data indicates that most organizations that are involved in translation do perform terminology work. Over 80 percent of respondents are active either globally or in several countries.

Forty-five percent of these respondents rarely or never use any type of language processing tool. This compares to 40% of organizations that perform terminology work (Survey 2), and 5% of organizations that have a termbase (Survey 3). Thus, organizations that do not perform terminology work are also more likely to not utilize other language tools compared to organizations that do perform terminology work.

The size of the organization (number of employees) does not appear to be a factor in the organization's decision not to perform terminology work. Large, medium, and small organizations are more or less equally represented in this group.

The reasons why these respondents indicated that they do not perform terminology work are, in order of frequency:

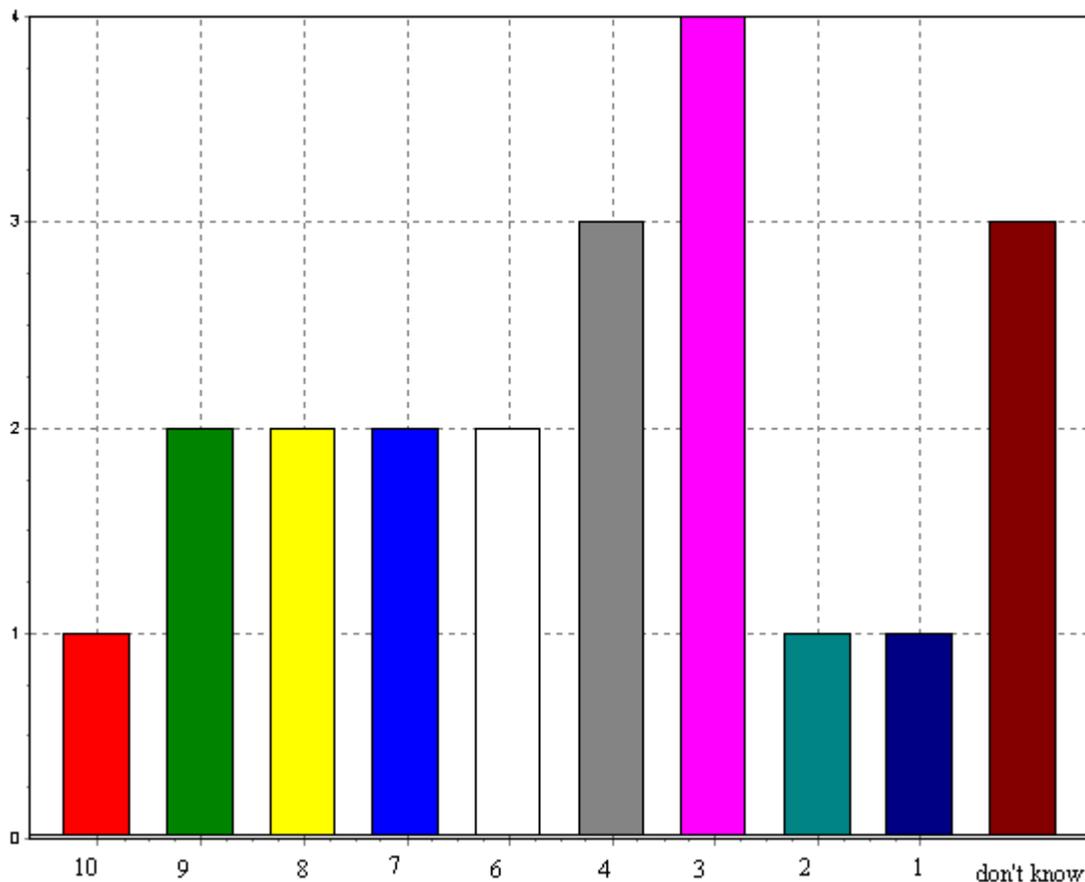
1. There is no perceived need.
2. Terminology management is not part of the organization's core business.

3. There are insufficient terms to manage.
About 20 percent of respondents felt that a lack of resources, knowledge, or skills played a role in their decision to not perform terminology work.

Nearly 60 percent of all respondents felt that performing terminology work would not significantly benefit their organization, or they did not know if there would be any benefits. Only 40 percent saw any significant benefits from performing terminology work. However, if you examine the responses of only those organizations that operate globally or in several countries, the percentage of respondents that see significant benefits rises to 50 percent.

Figure 1 - Rating of anticipated benefits if terminology management was performed

(1 = no benefit, 10 = very beneficial)



Observations

Only half of the 82 percent of respondents who are active in multiple linguistic markets see any benefits in managing terminology. Many respondents view terminology management, and language processing in general, as outside of their core business. Some

of these respondents outsource their terminology work, passing on the responsibility to a third party.

From the comments received, this group seems to view terminology management as an activity performed only by translators for translation purposes. There is a clear need to raise awareness that terminology is a key knowledge asset for a company. Too few industries that are not directly involved in translation or other language management activities are aware of the benefits of managing terminology in-house as a means of quality control. They are unaware of the full scope of terminology management, and that it can be applied even only to the source language to raise the quality of information assets, increase the potential for knowledge management, and make the company localization-ready. These industries might benefit from being told the success stories of other companies that have practiced terminology management, such as in the automotive and telecommunications sectors.

Respondents in this group who do see the benefits of terminology management most likely are the same ones who reported that they do not do it because of a lack of resources and skills.

Findings – Survey 2

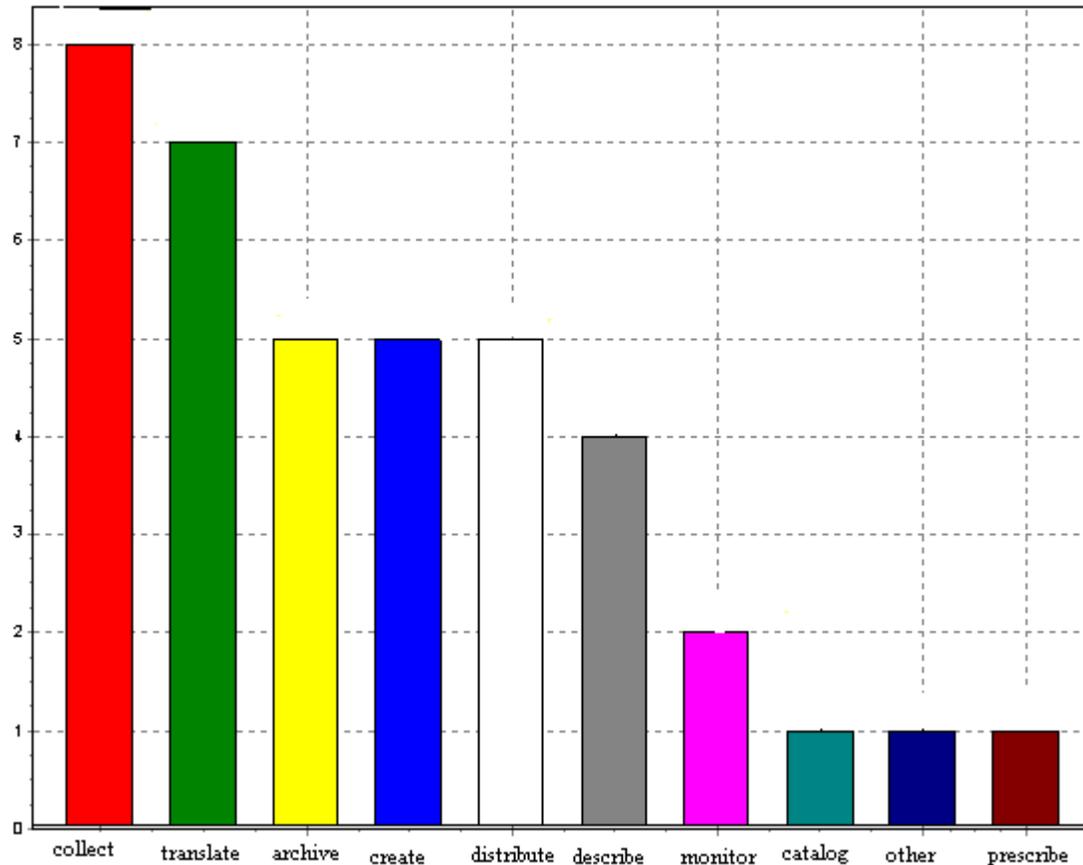
Survey 2 was completed by organizations that perform terminology work but do not maintain a termbase. Only ten responses were received for this survey, while 42 responses were received for Survey 3. This indicates that most organizations that perform terminology work also maintain a termbase.

The main goal of this survey was to learn about the nature and needs of organizations that perform terminology work without using a termbase, and why they do not have a termbase.

Half of the respondents are in the translation industry. Eighty percent are active globally or in more than one country. Sixty percent are frequent users of language processing tools. The majority of respondents represent small organizations, with 60 percent having less than ten employees.

The most common terminology activities performed are term collection and translation. Eighty percent of translation service respondents collect and store terms as an activity separate from the actual translation. Distributing and creating terms are less commonly performed by all respondents. Describing terms (creating definitions, etc.) is even less of a priority; only one of the translation respondents performs this task. Monitoring term usage, prescribing terms, and cataloging terms are performed by only ten percent of respondents.

Figure 2 - Terminology tasks performed by organizations that do not have a termbase.



Half of the respondents use a translation memory system, and of those, 90 percent are in the translation industry. Only one respondent uses a tool to extract terms or contexts, and only one respondent uses a translation workflow system.

The most frequently cited advantages of using language processing tools are increased speed and consistency.

The task that is most frequently performed manually is term identification and extraction.

Processes and resources that are rated as effective for translating terms are:

- Giving translators access to a bilingual termbase
- Having topic experts available to answer questions
- Having a translation revisor to check terminology

Having translators merely prepare their own bilingual dictionaries during translation was rated as less effective for translating terms.

Among all respondents, respecting industry standard terminology was rated as the most important criterion when selecting terms, followed by the need to keep the same terms throughout the life of a product or project (historical use). Linguistic criteria (derivability, etc.) are least important. However, when isolating the responses of the translation service providers, the results are slightly different. These respondents rated historical use of terms, industry standards, and company standards as most important, and national standards least important. Translation criteria (cultural neutrality, transparency, etc.) and linguistic criteria are considered more important for term selection by translation service providers than by other respondents.

Unlike the respondents of Survey 1, who do not perform terminology work because they do not see the strategic need, the most frequently cited reason that these respondents do not maintain a termbase is a lack of skills and knowledge. All respondents indicated that having a termbase would benefit their organization, with 70 percent of them estimating that it would provide significant benefits.

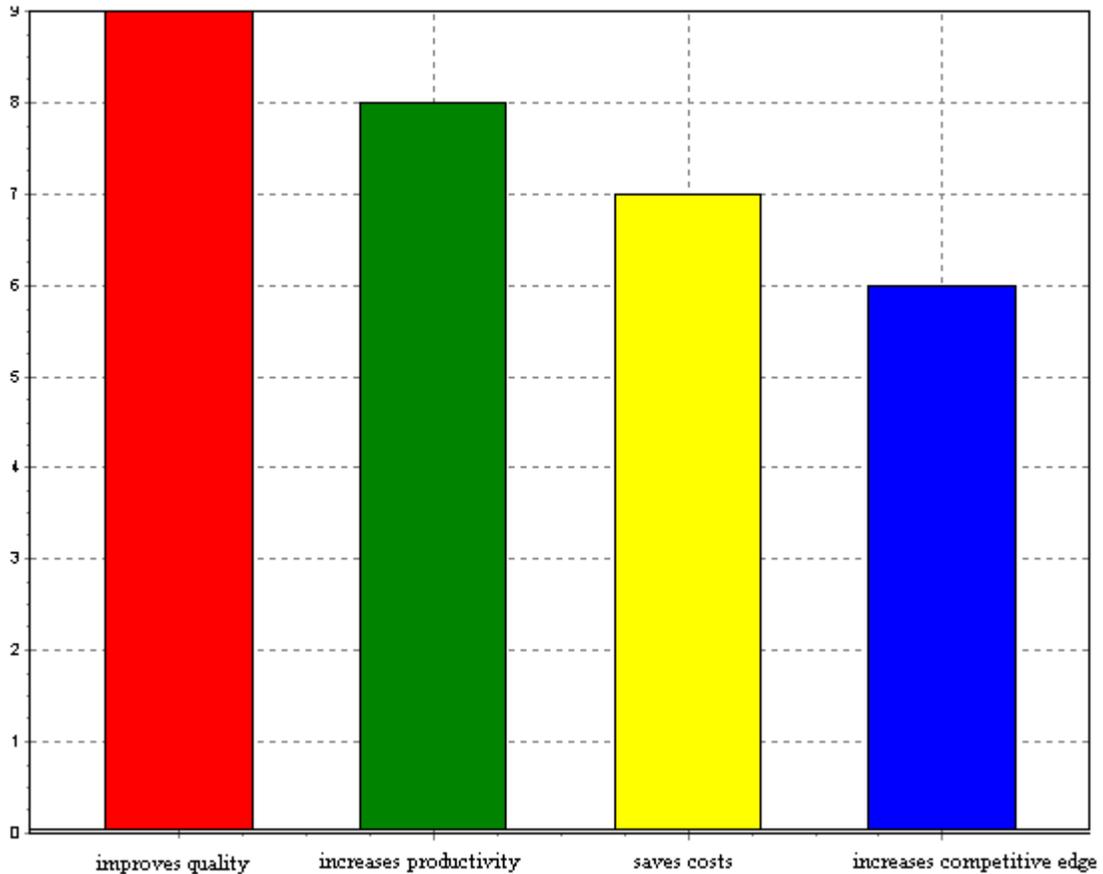
Seventy percent of respondents felt that industry glossaries are important for conducting terminology research, and 60 percent also chose ISO glossaries. Half the respondents chose dictionaries, reference texts, and standards as important reference material.

Seventy percent of respondents share terminology within their organization, while only one respondent (10 percent) makes it available to the general public. Increased consistency is cited as the most important benefit of sharing terminology. Overall, 60 percent of respondents do not need to convert their terminology data to any different format when sharing terminology.

Enabling data exchange was cited as the most urgent requirement for improving terminology work, with education and knowledge exchange a close second. The need to raise awareness among managers and clients about the need for terminology management was a recurring comment.

Ninety percent of respondents claim that terminology work improves quality, 80 percent claim that it increases productivity, and 70 percent estimate that it saves costs.

Figure 3 - Benefits of managing terminology for organizations that do not have a termbase.



Observations

This group, dominated by small translation companies, appears driven by immediate production needs to perform translations and consequently gives little attention to proactive language management. Comments suggest that these respondents must deal with multiple different clients and small jobs that do not warrant a long-term terminology plan. They appear to take few measures to ensure consistency in terminology. Furthermore, as small companies, they have less resources to invest in a termbase.

The fact that 4 out of 5 translation service respondents use a translation memory system, while none in this group maintain a termbase, indicates that translation memory systems have matured and gained widespread acceptance among smaller translation companies while terminology systems are lagging behind.

Given that term collection is one of the most frequent tasks performed, and is also usually performed manually, the fact that only one respondent in 10 uses a term extraction tool indicates that automating term collection would significantly increase productivity.

It is interesting to note that the most urgent requirement is to enable data exchange. In the context of the comments received, these respondents were most likely referring to sharing of resources such as industry glossaries. Since these respondents have limited resources of their own to perform terminology research, they would clearly be in a better position to provide high-quality terminology to their customers if more industry glossaries were available.

These respondents overwhelmingly agree that having a termbase would benefit their organization and the main reason that they do not have one is a lack of skills and knowledge. This contrasts with the respondents of Survey 1, the majority of whom did not see the need for a termbase. They also agree that performing terminology work generally improves quality, increases productivity, and reduces costs. They clearly see terminology work as more than simply having individual translators maintain their own bilingual dictionaries.

Findings – Survey 3

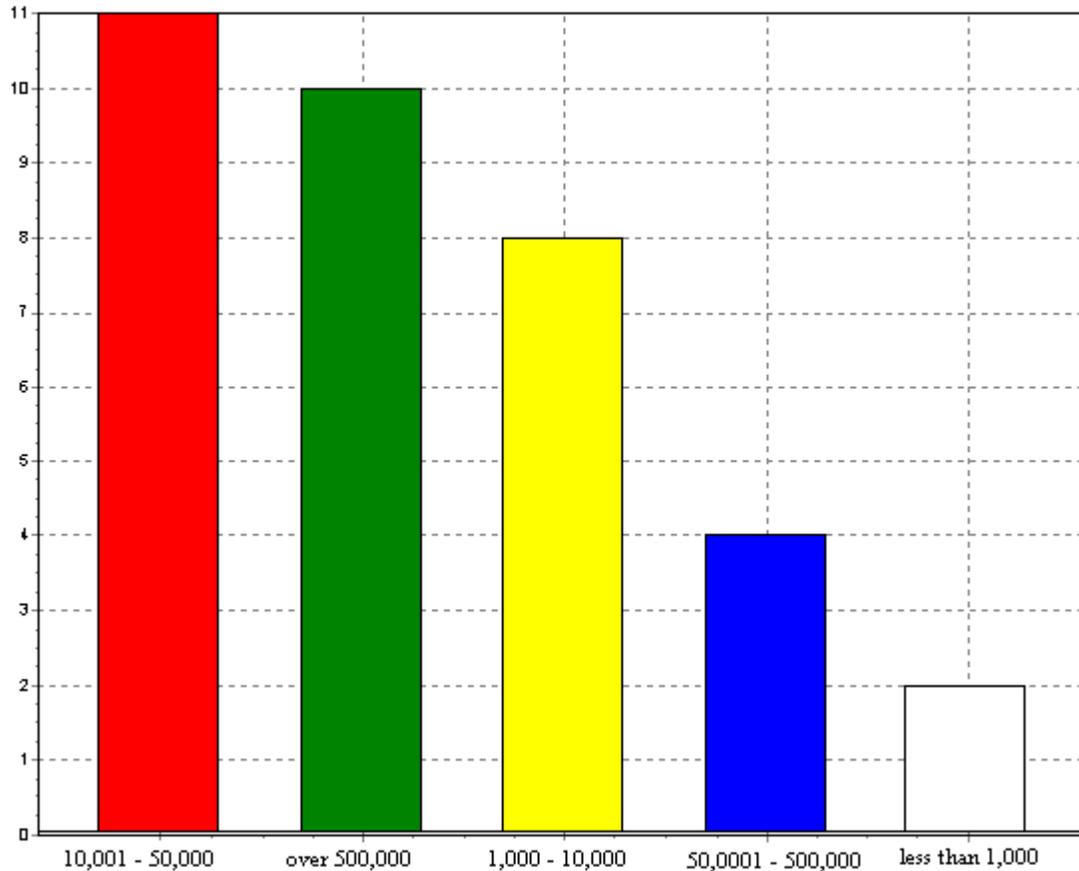
Survey 3 was completed by organizations that perform terminology work and also maintain a termbase. Forty three responses were received for this survey.

The main goal of this survey was to learn about the nature of industry termbases and how they support localization, as well as the needs of organizations that have termbases.

Thirty-seven percent of respondents are in the translation/localization industry, while 25 percent are software development companies (including developers of localization tools). The remaining 37 percent represent government, industry (telecommunications, medical), and academic institutions.

The termbases tend to be medium to large in size, with 30 percent of them containing more than 500,000 terms, 12 percent containing 50,000 to 500,000 terms, 32 percent containing 10,000 to 50,000 terms, and 23 percent containing 1,000 to 10,000 terms. Only six percent of the termbases contain less than 1,000 terms.

Figure 4 - Number of terms in the termbases



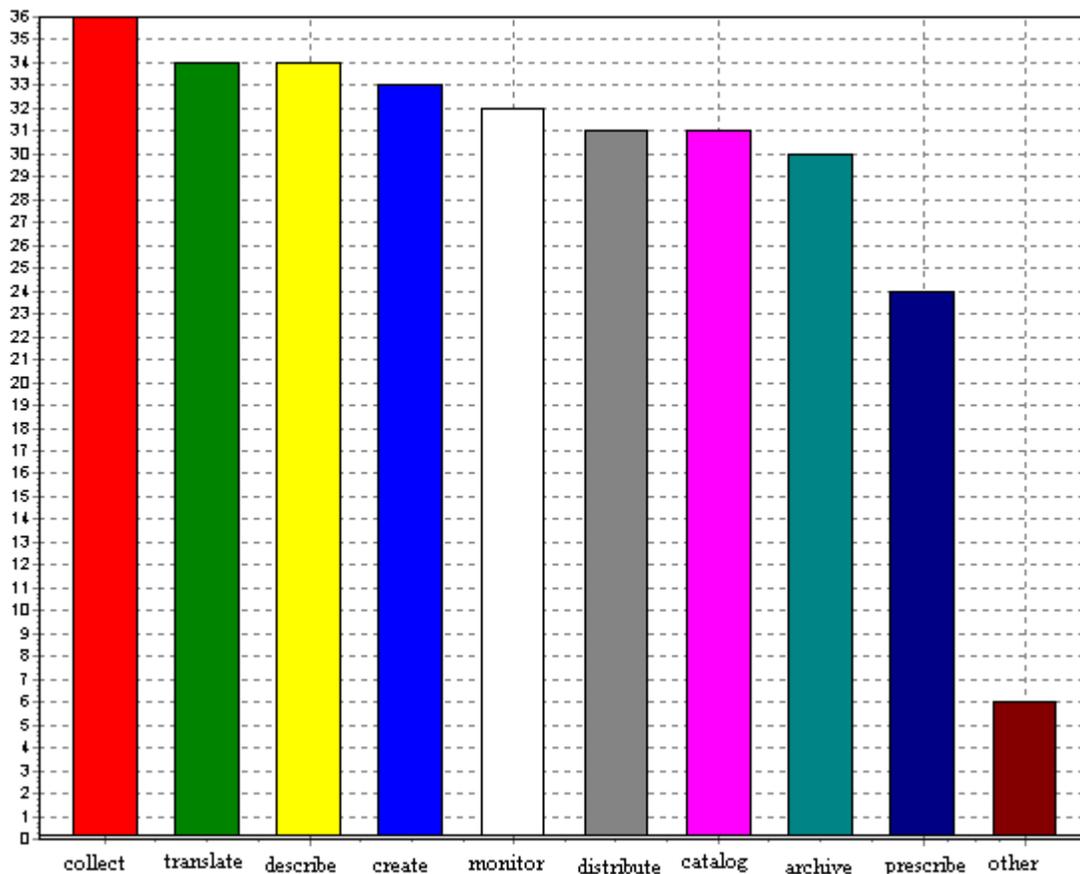
Almost 90 percent of respondents are active globally or in more than one country. Sixty five percent are frequent users of language processing tools while 30 percent develop language processing tools. Only five percent rarely or never use language processing tools (compared to 45 percent and 40 percent in Survey 1 and 2 respectively), indicating that organizations that maintain a termbase are also more likely to use other language processing tools than those that do not.

Seventy-five percent of respondents represent organizations that employ more than 100 people, while only 5 percent employ less than 10 people. This contrasts to the respondents of Survey 2, the majority of whom employ less than 10 people. This indicates that larger organizations are more likely to have the resources to develop a termbase.

As reported in Survey 2, collecting terms is the most frequently performed terminology activity, with 85 percent of respondents indicating that they do this task. Eighty percent also indicated that they translate terms. However, in contrast to the activities reported in Survey 2, these respondents also very frequently perform more complex terminology tasks, with 80 percent describing terms, monitoring term usage, and even creating terms.

Seventy five percent of respondents also distribute, archive, and catalog terms, and nearly 60 percent prescribe terms. Additional tasks were reported relating to more sophisticated concept management, such as researching concepts, aligning concepts across languages, eliminating synonyms, and standardizing terms.

Figure 5 - Terminology tasks performed by organizations that have a termbase



Almost 90 percent of respondents use a terminology management system, and 80 percent use a translation memory system, indicating that these two types of language processing tools are almost equally common among these respondents. Unlike the respondents of Survey 2, however, many of these respondents also use terminology format converters (70 percent), term extractors (60 percent), and consistency-checking tools (50 percent), and workflow management tools (44 percent). A significant number also use a wide range of other tools for related activities such as for coordinating terminology reviews, extracting contexts, and performing machine translation. These findings indicate that organizations that have a termbase are sophisticated users of language processing tools in general.

The most frequently reported advantage of using language processing tools is increased consistency and quality followed closely by time savings. The most frequently reported

disadvantage is a lack of compatibility and interoperability between the various tools, followed by investment costs and incompatible file formats. Several respondents reported that the use of translation memory tools leads to translation errors and reduced stylistic quality of translations due to the inability of the translation memory system to interpret contexts.

As reported in Survey 2, the terminology-related task that is most frequently performed manually is term and context extraction. Workflow management and quality control (checking for term consistency) are also frequently performed manually. Other manual tasks that were reported include tracking terminology reviews and some housekeeping tasks for the termbase, such as finding duplicate terms, finding incomplete records, and determining areas that require research. One respondent noted that importing and exporting terminology data in termbases requires some manual effort due to a lack of standard exchange formats.

All respondents felt that giving translators access to a bilingual termbase is an effective strategy for translating terms. Nearly 40 percent reported that separating terms out of the corpus before they are translated is beneficial for their translation process, while 35 percent felt it is beneficial to translate them at the same time as other parts of the corpus. Half of the respondents reported that having a topic expert available to answer translators' questions, and having translators compile their own bilingual dictionaries, are other effective strategies.

Seventy percent of respondents felt that checking terminology when reviewing translations is an effective strategy for translating terms, but virtually all respondents that actually provide translation/localization services reported that this is very effective. Sixty percent of all respondents felt that it is effective to review terminology during final testing and quality assessment activities, whereas 80 percent of the translation/localization services respondents felt that this is effective. Other effective strategies mentioned are to have terminology approved by subject matter specialists and authorized by the client *before* translation begins.

The criteria most frequently applied when selecting terms for use are adherence to company/ organization standards (85 percent of respondents), followed closely by historical use of the term in the context and industry standards (75 to 80 percent). Respect for international standards, national standards, translation criteria (cultural neutrality, etc.) and linguistic criteria (derivability, etc.) are deemed less important overall (60 to 65 percent). Translation criteria (cultural sensitivity, etc.) and linguistic criteria were, however, rated higher by translation respondents than other groups.

The advantages reported of applying these term selection criteria include:

- Higher quality of terms
- Improved usability, readability, and translatability
- Higher degree of recyclability of terms
- Reduced translation costs

- Reduced localization efforts
- Greater acceptance by target market

It was, however, also noted that over-applying the historical use criterion can have negative effects:

- Resistance to change incorrect terms
- Negative impact on style
- Reduced emphasis on usability

Term selection criteria can vary according to the client's wishes. Some clients prefer industry standard terminology while others want company-specific terms. This presents challenges to translation service providers who adopt various strategies such as documenting client preferences in their termbase or creating separate termbases for each client.

One respondent repeatedly commented that inconsistent terminology reduces the effectiveness of translation memory and machine translation tools.

Sixty-five percent of all respondents, and 90 percent of translation/localization service providers, employ some means to ensure that the same translation is used consistently for a given source term. The most frequently cited means to achieve this is by using a translation memory system, followed by manually checking and then by using a termbase. Only 10 percent of respondents stated that they used an automated terminology checking tool, and whether they were referring to a tool that actually checks for consistency at the term level is unclear.

The resource that was selected by the most respondents overall (85 percent) as useful for researching terminology is traditional dictionaries, followed in sequence by industry glossaries, reference texts, and glossaries from standards bodies. However, translation respondents value dictionaries, industry glossaries, and reference texts equally as most important. Respondents also mentioned external terminology databases and customer glossaries as important resources.

Resources that respondents indicated they do not currently use but would find useful include:

- Cross-industry glossaries
- External termbases
- ISO glossaries
- Standards about methods
- Corpuses

The reasons indicated for not using these resources are cost (of ISO glossaries, standards, and some termbases), unavailability (of cross-industry glossaries), and lack of time (to research corpuses).

Eighty five percent of respondents share terminology within their organization, and 25 percent make their terminology public and/or share their terminology with external organizations, including sometimes competitors (compared to 70 percent and 10 percent, respectively, in Survey 2). Twenty-five percent also share their terminology with customers, partners, or vendors. Many respondents do not share customer-specific terminology for confidentiality reasons. Increased consistency and quality is the most frequently cited benefit of sharing terminology, followed by knowledge reuse and time savings. Twelve percent of respondents indicated that sharing terminology is a branding strategy that helps position the company as an industry leader.

Overall, 80 percent of respondents need to convert their terminology data to a different format when sharing terminology (compared to only 40 percent in Survey 2). The formats used, from most to least responses, are:

1. Comma-separated or tab separated or delimited ASCII
2. Spreadsheets
3. MultiTerm
4. Unspecified encoded text (such as HTML)
5. XML
6. Word processing files
7. SGML

The responses to the question “What are the main challenges in your organization's terminology work?” can be categorized in the following themes, from most frequent to least frequent:

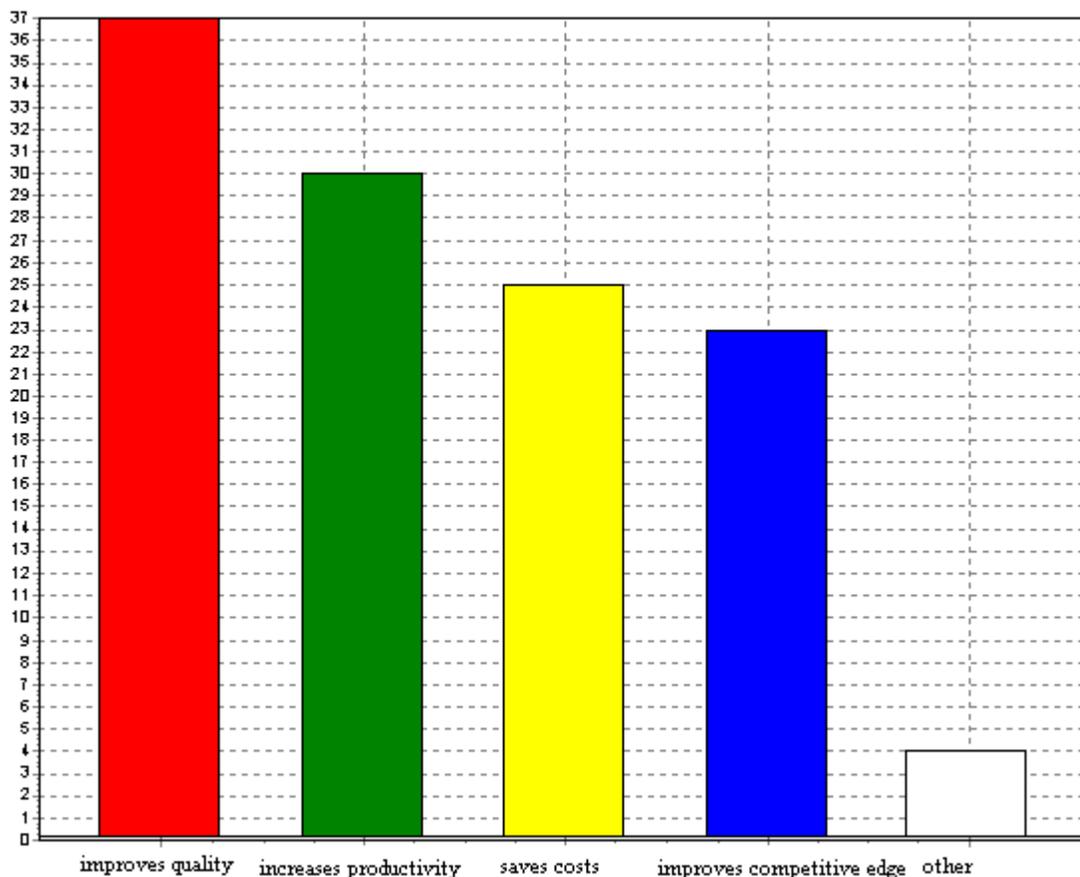
1. Simply coping with limited budgets and resources and little time
2. Raising awareness of the need for terminology management among clients, corporate executives, and other employees
3. Harmonizing terms and concepts
4. The following responses have equal weight in position 4:
 - Dealing with new concepts and terms
 - Dealing with a lack of skills
 - Dealing with inadequate tools or non-integrated tools
 - Validating and revising terminology
 - Enforcing use of terminology
 - Managing the workflow
5. The following responses have equal weight in position 5:
 - Estimating return on investment
 - Categorizing terminology to suit audience needs (such as by using subject fields)
 - Changing bad legacy terms
6. Dealing with a resistance to share terminology
7. Supporting diverse user needs
8. Establishing company-wide processes.

Half the respondents indicated that providing education seminars for staff, and facilitating data exchange, are urgently required in order to improve terminology work, and almost half felt that increasing knowledge of industry standards is also urgently needed. Thirty percent felt that fostering knowledge exchange with other organizations is needed. Others reiterated the need for better tools and increased resources (staff, budgets).

Over forty percent of respondents felt that obtaining resources for terminology work is more difficult than for other initiatives, while 30 percent felt that this requires the same effort, and only 15 percent felt it is easier.

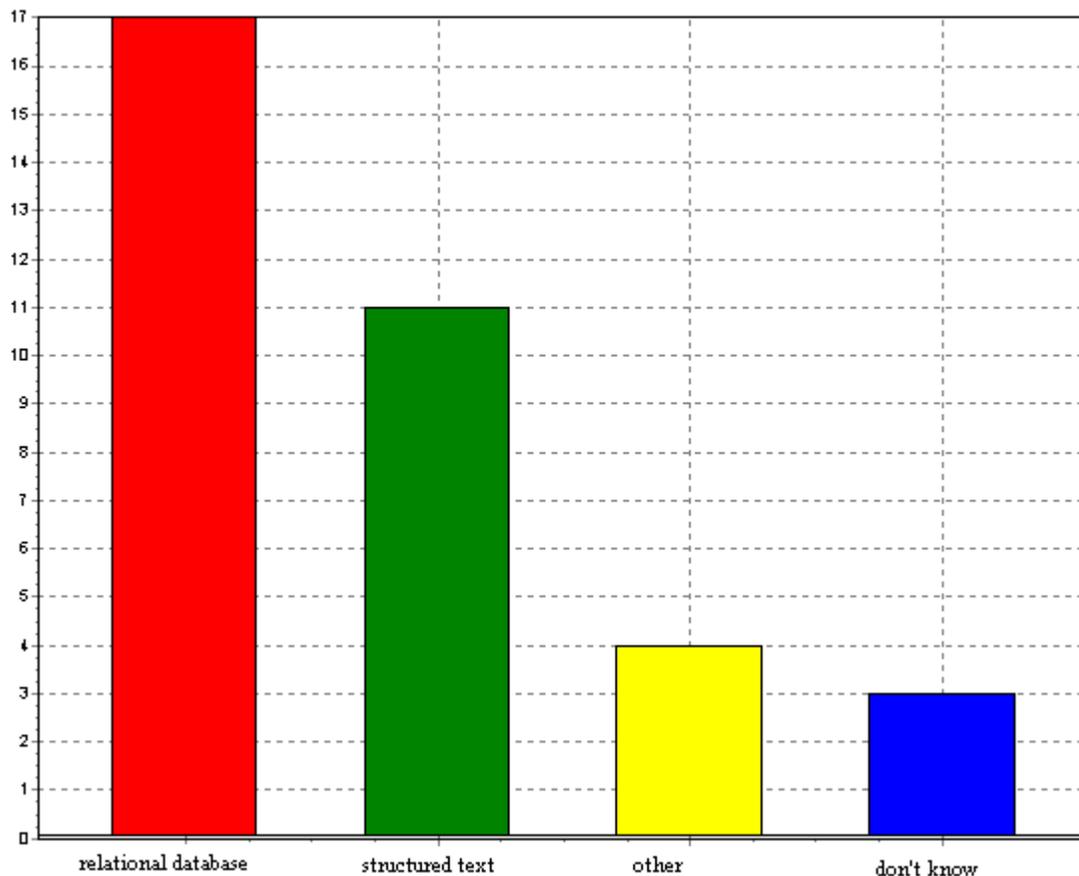
Ninety percent of respondents claim that terminology work is justified by improved quality, 73 percent by increased productivity, and over 60 percent by cost savings. These results are very similar to those obtained in Survey 2. However, over 55 percent of the respondents of Survey 3 also feel that terminology work improves competitive edge. One respondent claimed that terminology work actually increases revenues. It was also mentioned that terminology can be a resource to reduce initial development costs for new projects and it can supply data for other language tools.

Figure 6 - Benefits of managing terminology for organizations that have a termbase



Over half of the respondents use a relational database as the core technology for their termbase, while 35 percent use a structured text system (such as MultiTerm), and only one respondent (3 percent) uses a “flat file” system. Almost 10 percent of respondents did not know what core technology is used. A number of respondents revealed that their reason for selecting a particular technology or tool was not founded on strategic or long-term need; instead, they gave reasons such as “it was there,” “it was first to market,” or “it is also used by our vendors.” Some of these respondents indicated that as a result of this lack of planning they were now preparing to switch to a different core technology.

Figure 7 - Core technology of the termbase



Respondents reported a wide range of business needs that are not currently being met by their termbase. Recurring problems are:

1. Lack of a Web interface
2. Lack of integration with other tools such as the translation memory system, controlled English system, and machine translation system
3. Poor management of subsets, such as to manage terminology for different clients in the same database

4. Lack of conversion routines to formats such as XML
5. Lack of statistics reporting features
6. Lack of housekeeping functions and features to enforce quality control during data entry.

The respondents' rating of data categories for the source language and target language is shown below, arranged from most important to least important. The overall rating was determined by counting the number of respondents who had rated each category a value of 7 or above out of a maximum rating of 10.

Source language:

1. Term
2. Context
3. Source of term
4. Definition
5. Term quality indicator
6. Product identifier
7. Subject field
8. Usage note
9. Source of context
10. Source of definition
11. Other grammatical data
12. Part of speech
13. Concept relations
14. Other miscellaneous data categories (customer ID, subset values, variants, transfer comment, user ids, date stamps, MT-ready flag)
15. Regional usage indicator
16. Usage register

Target language:

1. Term
2. Context
3. Source of term
4. Term quality indicator
5. Definition
6. Subject field
7. Product identifier
8. Regional usage indicator
9. Source of context
10. Part of speech
11. Usage note
12. Other grammatical information
13. Source of definition
14. Usage register
15. Other miscellaneous data categories (customer ID, subset, variants, transfer comment, user ids, date stamps)

16. Concept relations

It is worth noting for both the source language and the target language:

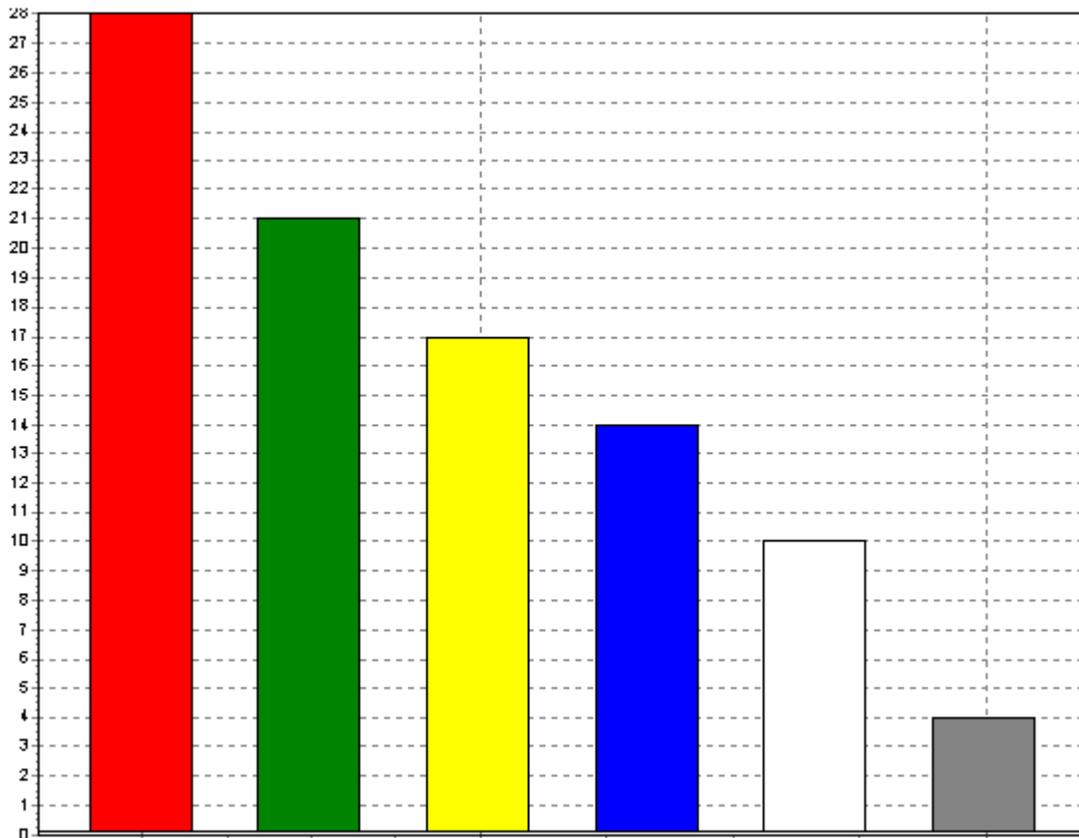
- The context is more important than the definition.
- Recording the source of the term is very important.
- Subject fields and product identifiers are also important data categories.

The main differences that can be noted between the source language and target language are:

- It is more important to define the source term than the target term.
- It is much more important to record the regional usage properties of the target term than the source term.
- It is more important to record concept relations in the source language than the target language.

Respondents' termbases are typically focused on the organization's sphere of activity. Thirty five percent of termbases also include terms from specialized domains outside of the organization's sphere of activity, but only 25 percent include words from the general lexicon. It is more common to include all terms found in use rather than restrict the termbase to only approved terms. (The fact that the term quality indicator was rated as an important data category supports this broader inclusion approach by making it possible to indicate which terms are approved and which are not.)

Figure 8 - Types of terms included in termbases



Legend

- Terms from specialized domains within the organization's sphere of activity
- All terms found in use.
- Only approved terms.
- Terms from specialized domains beyond the organization's sphere of activity.
- Words from the general lexicon.
- Other.

Respondents commented that term inclusion criteria can vary by project, by client, and by the aims of terminology management. For example, a termbase that is intended to disseminate standardized terms might only include standardized terms.

Reasons cited for excluding words from the general lexicon include:

- Reduce work effort
- Make the termbase more useful to future users

- Focus on translators' needs
- Keep the termbase from becoming too large

An interesting justification expressed for including all terms in use was that it helps users to understand a new term if they see what terms it replaces.

The methods of accessing the terminology in a termbase were rated as follows, from most important to least important:

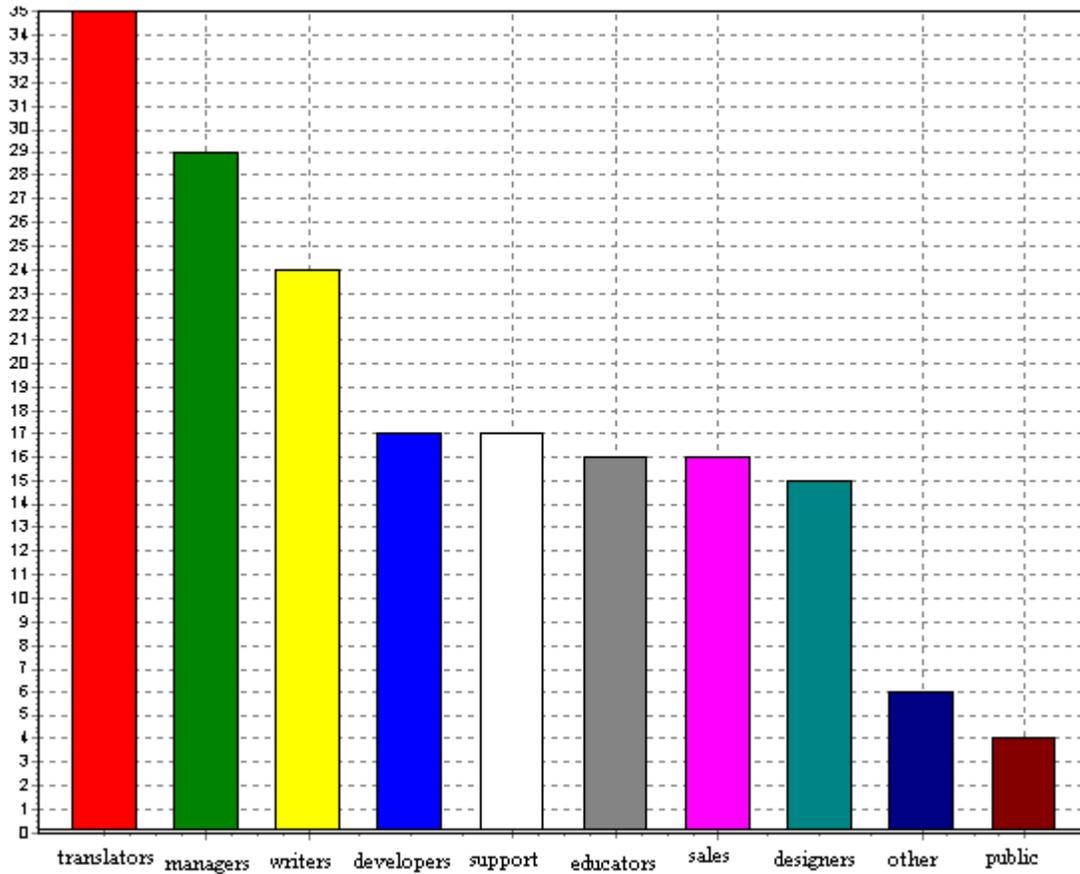
1. Through a program that users install on their workstations
2. Through a Web site
3. Through other applications, such as text editors
4. Through a dial-in connection
5. Through a CD-ROM

The advantages noted for a Web site are that it provides universal real-time access and it is always up to date. The main disadvantage is performance (Internet speed). Locally-installed programs are more flexible in that they can be customized to address different user requirements and they can have better performance. The disadvantages of local programs are that the data may be out of synch with other copies of the termbase, they require frequent extractions and updates, and they can be costly. (This can also be said of CD-ROMs.)

Numerous respondents commented that a variety of access methods are necessary to meet different user group needs. The preferred method for general users is a Web site, whereas translators are better served by formats specifically tailored to the translation activity in a program that can be integrated into their work environment.

Eighty-five percent of the termbases are used by translators, 70 percent by managers, and 60 percent by writers. Less than 40 percent of the termbases are used by support staff, product developers, sales staff, educators, and product designers. Only 10 percent of the termbases are used by the general public.

Figure 9 - Who uses the termbase



The rate of use of output from the termbases in other tools is low. Twenty-five percent or less of respondents indicated that the data from their termbase is used in another tool.

Almost 30 percent of respondents could not estimate the cost of maintaining their termbase. Twenty three percent of the termbases cost less than one dollar U.S. per term to maintain yearly, 25 percent cost between one and five dollars U.S. per term, and 20 percent cost over five dollars U.S. per term.

Virtually none of the respondents could estimate the cost savings realized by their organization as a result of using the termbase. Only one respondent ventured a guess: two to three dollars per term per year.

Observations

The fact that respondents of Survey 3 perform more complex terminology tasks than the respondents of Survey 2 suggests that some organizations do not manage terminology proactively simply because they do not have appropriate tools to do so. The finding that describing terms and concepts is considered by this group to be as important as

translating them reveals the importance of the more sophisticated terminology tasks in support of translation. The fact that the majority of respondents monitor term usage and prescribe terms suggests that having a termbase enables an organization to proactively manage terminology.

The fact that 60 percent of these respondents have found or developed a term extractor tool indicates that this group endeavors to use tools to automate labor-intensive processes as much as possible. These respondents have resources and experience in this area that could benefit smaller translation companies (Survey 2).

Improving the interoperability of language processing tools is urgently required to increase productivity.

Terminology extraction and workflow management are two tasks that would significantly benefit from automation.

The differences in responses regarding the effectiveness and necessity of reviewing terminology as part of the translation process between those who provide translation/localization services and those who do not indicate that translators and localizers have a higher awareness of terminology problems that arise during translation than other respondents such as clients of these services.

The fact that the majority of respondents rely on their translation memory system to establish terminology consistency, complemented by a manual checking process, and possibly casual checking of a termbase, suggests that there is a lack of a recognized tool to perform this task. Since translation memory systems typically function on the phrase level not the term level, they are not specifically designed for verifying terminology consistency. Manually checking is of course error-prone, and a termbase can only help if the user initiates the check. There seems to be a need for tools that can more efficiently check for terminology consistency, such as between separate translator files.

The fact that translation respondents rated reference texts as important as glossaries and dictionaries for researching terminology indicates that traditional lexicological resources are insufficient on their own and that a corpus of authentic material provides essential information for translators and terminologists.

Organizations that maintain a termbase are more likely to share terminology than organizations that do not.

Some companies consider their in-house terminology to be confidential and prohibit their translation service providers from sharing this terminology. This appears to be in conflict with a desire by translation service providers to have access to cross-company industry terminology. Translation respondents expressed frustration that cross-industry terminology is not generally available and that they themselves are often restricted from sharing this terminology. On the other hand, it can sometimes be absolutely critical to share terminology across an industry. One respondent, representing a national defense

department, stated that sharing military terminology among allied countries is essential to permit interoperability among cooperating nations. These points suggest that consumers of translation services need to be educated about the benefits of sharing their terminology – translation services cannot be expected to use high-quality, standardized terminology in their translations if terminology continues to be considered confidential by industry.

The fact that most termbase owners need to convert terminology to different formats for sharing purposes, while few organizations that do not have a termbase do any data conversion, indicates that knowledge and experience in data formats and tools for data conversion are essential for termbase management. Organizations that are involved in standardizing terminology data formats, such as LISA and ISO, should establish close ties with termbase owners and developers of terminology management systems to promote adoption of these standards.

The way in which respondents answered the questions about the core technology of their termbase suggests that, for a significant number at least, their knowledge about the core technology is low. It is unfortunate to see that some organizations did not thoroughly research their needs and the available technologies before selecting their terminology management system, and that some of them, as a consequence, are faced with the difficult prospect of changing terminology systems after their termbase has already been established.

In the rating of data categories, it is interesting to note that recording concept relations was not considered very important. This should not be too surprising, as it is a well-known fact that most termbases are not truly concept-oriented. There may be several reasons for this:

- The terminology management system used does not efficiently manage concept relations. Several comments were made to this effect elsewhere in the survey.
- The focus on spontaneous translation activities rather than controlled language reduces terminology work to a lexicographic approach.
- There is a lack of awareness of the benefits of recording concept relations.

Comments were made that terminology consistency in the source language is important and needs to be improved, yet less than 40 percent of termbases are used by product developers and designers and other professionals who influence terminology use (aside from traditional users such as writers and translators). This reveals that termbases are being underutilized in the source language community as a means of increasing terminology consistency.

The inability of most respondents to estimate the maintenance costs of their termbase or the dollar value of benefits to the organization suggests that better administration records need to be kept and more studies need to be done on the economics of terminology management. Perhaps termbase administrators could benefit from taking some educational seminars in business management.

Overall Conclusions and Recommendations

This section summarizes some broad conclusions and recommendations that can be drawn from the three terminology surveys.

Non-Communications Industries

Globally active organizations whose core business is not communications-related (translation, localization, information management, etc.) are generally unaware of the benefits of performing terminology management in-house including:

- Better quality of information materials
- Reduced translation costs
- Increased brand image

They are also unaware that they can benefit from working more closely with their translation service providers in establishing acceptable terminology. This partnership simplifies translation (which reduces costs for the client) while improving quality.

Controlled English measures or tools can facilitate translation and localization. But in order to optimally support this process, the controlled English system must be supplied with standardized terminology. Thus it is apparent that globally active organizations, even if they outsource their translation and localization needs, have an interest in proactively managing their source language terminology.

If industries were less protectionist about terms, terminology consistency and standardization would increase.

Non-communications industries that want to initiate some in-house terminology management and facilitate correct terminology use by their translation service providers should consider establishing a termbase. They should thoroughly research their needs before selecting a terminology management system and use the successful experiences of other similar organizations as models of implementation.

Translation and Localization Services and Organizations That Have a Termbase

Given the evidence that a termbase is required to proactively manage terminology, which increases consistency and quality of information, enhances productivity, and reduces costs, and the additional advantages that a termbase provides, such as the increased ability to exchange terminology and to provide it in various output formats, translation and localization services that do not have a termbase would be well advised to get one. But they should not adopt the first terminology management system that comes along. To avoid having to switch systems later, they should thoroughly evaluate all available systems and their ability to meet specific needs. Small companies may have resisted

developing a termbase in the past because of the investment costs. But the advantages of termbases shown in Survey 3 might now justify the investment for companies that want to remain competitive.

Educating clients of translation services about the importance of terminology tasks, such as pre-establishing standard terms, allowing time for revision, and permitting terminology changes during final testing, may help to improve the level of client/supplier collaboration to achieve the highest quality terminology. Suppliers of translation services should be sensitive to the fact that many of their clients may be unaware of the processes that are necessary and the cycles required to produce translations with consistent, accurate terminology.

Establishing the terminology needs of clients before translation begins and before inputting terms into a termbase will help in planning terminology work.

Organizations that already have a termbase could consider offering educational seminars on the following topics for their terminologists:

- Technical skills, such as those required to create custom conversion routines
- Standard principles and methods of terminology management
- The base technology used by the terminology management system, and how to optimize the use of this technology
- Business management skills, particularly:
 - Financial skills to monitor maintenance costs and to determine ROI
 - Time management
 - Project management
 - Strategy-building skills – developing a long term plan, obtaining buy-in from upper level management, etc.

The following strategies for termbases were commonly followed or identified as important, so they can be assumed to be effective for terminology management in most instances:

- Develop term inclusion criteria that meet the specific needs of your target users. Generally, it is best to exclude words from the general lexicon from the termbase.
- If you cannot include a definition, try to at least include a context.
- Plan the categories you will need to organize terms into groups required by your users and implement them in the termbase as subject areas, subsets, product identifiers, and so forth.
- Include all terms in use and not just “approved” terms. (If you only include approved terms, a user who does not find a term does not know if it is a prohibited term or if it has simply not yet been included in the termbase.) To guide usage, label terms appropriately to indicate whether they are approved, prohibited, etc.

The benefits of having a termbase would increase if more people were encouraged to use it. Termbases should be more promoted overall, but especially for non traditional users

such as product developers, designers, educators, sales staff, and so forth. The responsibility for terminology quality and consistency starts with the very first document written for a product or project, which often originates from staff other than writers and translators.

Developers of Translation and Localization Tools

Developers of translation and localization tools should focus on the top requirements stated by their customers:

- Improved integration and interoperability of tools: translation memory systems, terminology management systems, style/spelling/terminology checkers, MT systems, distribution systems (Web interfaces, hardcopy outputs, etc.)
- Fully functional Web interfaces
- Improved statistical reporting functions
- Improved terminology housekeeping functions: identifying duplicate terms, incomplete records or incomplete domains, improving user access levels, etc.
- Improved functions to manage subsets of terminology, for projects, clients, etc.
- Efficient term and context collection tools
- Streamlined translation workflow management systems
- Tools specifically tailored to check for terminology consistency (as opposed to translation memory tools that function on longer segments)
- Fully implemented standard data exchange formats

Assuming that the need for managing concept relations will increase, especially as organizations consider using terminology in less traditional applications such as search engines, tools developers may want to consider how to improve the concept-management capabilities of terminology management systems.

Closing Remarks

Whether or not they have a termbase, respondents representing organizations directly involved in localization recognize the benefits of performing proactive terminology management. Most perceive terminology management as a broader activity than simply translating terms. Terminology begins at the source through such initiatives as controlled English and source language terminology monitoring, and continues through product localization and distribution in target markets, passing through a wide range of tools and formats. It involves various players such as product designers and developers, writers, and translators. Organizations that maintain a termbase do so because it enables them to manage this complex process most effectively. Yet staffing, budgets, recognition, adequate tools, professional training, clearly-defined processes, and the availability of terminology resources remain a challenge for many if not most people active in this field.

On the other hand, organizations that are customers of localization services or whose core business is not focused on communications or localization tend to ignore the benefits of terminology management and perceive it as a narrow activity of interest only to translators.

Raising awareness of the benefits of terminology management will help promote its acceptance as an important component of the localization process and thereby foster positive change in this area. It is hoped that this survey will help achieve that objective.

References

At the time this report was written, the three terminology surveys can be viewed at the following Web sites:

Survey 1

<http://websurveyor.net/wsb.dll/4551/LISATerminologySurvey1version1.htm>

Survey 2

<http://websurveyor.net/wsb.dll/4551/LISATerminologySurvey2version1.htm>

Survey 3

<http://websurveyor.net/wsb.dll/4551/LISATerminologySurvey3version1.htm>

If these Web sites become unavailable, you can get a copy of the surveys by e-mailing Kara Warburton (kara@ca.ibm.com).