

Managing Terminology for Translation Using Translation Environment Tools: Towards a Definition of Best Practices

by

Marta Gómez Palou Allard

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Translation Studies

Under the supervision of
Dr. Lynne Bowker and
Dr. Elizabeth Marshman

School of Translation and Interpretation
Faculty of Arts
University of Ottawa

© Marta Gómez Palou Allard, Ottawa, Canada, 2012

Abstract

Translation Environment Tools (TEnt's) became popular in the early 1990s as a partial solution for coping with ever-increasing translation demands and the decreasing number of translators available. TEnt's allow the creation of repositories of legacy translations (translation memories) and terminology (integrated termbases) used to identify repetition in new source texts and provide alternate translations, thereby reducing the need to translate the same information twice. While awareness of the important role of terminology in translation and documentation management has been on the rise, little research is available on best practices for building and using integrated termbases. The present research is a first step toward filling this gap and provides a set of guidelines on how best to optimize the design and use of integrated termbases.

Based on existing translation technology and terminology management literature, as well as our own experience, we propose that traditional terminology and terminography principles designed for stand-alone termbases should be adapted when an integrated termbase is created in order to take into account its unique characteristics: active term recognition, one-click insertion of equivalents into the target text and document pretranslation.

The proposed modifications to traditional principles cover a wide range of issues, including using record structures with fewer fields, adopting the TBX-Basic's record structure, classifying records by project or client, creating records based on equivalent pairs rather concepts in cases where synonyms exist, recording non-term units and multiple forms of a unit, and using translated documents as sources.

The overarching hypothesis and its associated concrete strategies were evaluated first against a survey of current practices in terminology management within TEnt's and later through a second survey that tested user acceptance of the strategies. The result is a set of

guidelines that describe best practices relating to design, content selection and information recording within integrated termbases that will be used for translation purposes. These guidelines will serve as a point of reference for new users of TEnTs, as an academic resource for translation technology educators, as a map of challenges in terminology management within TEnTs that translation software developers seek to resolve and, finally, as a springboard for further research on the optimization of integrated termbases for translation.

Résumé

Les environnements de traduction (ETs) se sont popularisés au début des années 1990 en réponse à l'augmentation de la demande pour les traductions et à la progressive diminution de traducteurs sur le marché du travail. Les ETs utilisent des collections de traductions antérieures (mémoires de traduction) et de terminologie (bases de terminologie intégrées) pour identifier des passages répétitifs dans les nouveaux textes source et proposer des options de traduction afin d'éviter la retraduction de textes. Quoique la reconnaissance de l'importance du rôle de la terminologie en traduction et gestion de la documentation soit à la hausse, il y a un manque évident de recherche sur les meilleures pratiques de création et d'utilisation des bases de terminologie intégrées. Ce projet de recherche se veut un premier pas pour combler ce manque en proposant une série de stratégies définies par l'utilisateur qui visent à optimiser la conception et l'utilisation des bases de terminologie intégrées.

À partir d'études sur les outils d'aide à la traduction et à la gestion de la terminologie, et de nos expériences personnelles, nous avons conçu une hypothèse d'ensemble qui soutient que l'on devrait adapter les principes terminologiques et terminographiques traditionnels aux caractéristiques uniques des bases de terminologie intégrées : la détection automatique de termes et l'insertion conviviale des équivalents dans le texte cible.

Les modifications aux principes traditionnels comprennent l'utilisation d'un moindre nombre de champs dans la fiche terminologique, l'implantation de la structure de fiche proposée dans TBX-Basic, la classification des fiches par projet ou client, la création de fiches par paires d'équivalents et non par concept en cas de synonymie, la collection d'unités autres que les termes et de formes d'unités autres que la forme de base, ainsi que l'utilisation de textes traduits comme sources d'équivalents.

L'hypothèse d'ensemble et les stratégies pour l'exécuter ont été évaluées d'abord en

sondant les pratiques actuelles de gestion de la terminologie avec les ETs et ensuite à l'aide d'un deuxième sondage qui demandait aux participants de juger l'utilité de chaque stratégie.

Le résultat final est une série de conseils qui décrivent les meilleures pratiques de conception, sélection de contenu et saisie d'information dans des bases de terminologie intégrées destinées à être utilisées à des fins de traduction. Ces conseils serviront comme référence aux nouveaux utilisateurs d'ETs, ressource aux éducateurs des technologies de la traduction, indicateur des défis de la gestion terminologique au sein des ETs que les développeurs d'outils d'aide à la traduction pourraient se proposer de résoudre, et, enfin, tremplin pour des recherches futures sur l'optimisation des bases de terminologie intégrées destinées à la traduction.

Acknowledgements

First and foremost, I want to thank my thesis directors, Lynne Bowker and Elizabeth Marshman, for their guidance and support throughout this project. They have been an invaluable point of reference, both academically and personally. I have deeply appreciated their company along the sinuous path of writing a doctoral thesis. In spite of their demanding academic schedules and their own personal lives, they have always been there— with a constant smile and words of encouragement – to act as expert advisors, testers, sounding boards, editors, life counsellors, and, ultimately, friends.

I gratefully acknowledge the funding sources that alleviated the financial burden of embarking on a Ph.D. program. I was honoured to receive a University of Ottawa *Bourse de la francophonie* for the first four years of my program, as well as ATIO's Excellence scholarship in my first year.

This research is based in part on two on-line surveys that were distributed in 2009 and 2011. A total of 327 members of the translation community participated in the surveys. Without this vibrant response from my fellow translators, terminologists, project managers and CAT-tool specialists, this thesis would not have been possible. Apart from all my colleagues and acquaintances who were kind enough to help me distribute these surveys, two industry experts broadcasted the existence of the surveys and had a significant impact on participation: Jost Zetzsche reported on each of the surveys in his biweekly translation technology newsletter, *The Tool Kit*, and Barbara Inge Karsch joined him for the second survey on her on-line blog, *BIK Terminology*. I hope to be able to repay the community's generosity with the contribution this research brings to our field.

Special thanks must go also to the colleagues who participated in each survey's pilot testing and provided comments on the formulation of questions, skip logic and the overall flow of each survey. Their feedback helped to improve the surveys and contributed to their success.

The Translation Bureau kindly shared with me and my thesis directors two unpublished research studies undertaken by Guy Champagne in 2004: *Portrait of Terminology in Canada* and *The Economic Value of Terminology: An Exploratory Study*. The findings in both studies enriched this project's literature review.

The University of Ottawa not only provided me with the academic resources to further my education but also introduced me to the friends who supported me (in all the senses of the verb) during these past years. Our coffees, suppers, trips, *peeps'* nights, board games, house moves, life celebrations and, especially, our laughs infused these years with fun and made them so much more enjoyable. I cannot name all of them, but I would especially like to recognize Kim, Julian and María Sierra, who walked by my side from the first steps of this thesis. The passionate discussions on terminology with Kim managed to revive my interest in the project even through the driest of the sections and, after her inspiring conversations and delicious meals, I was always ready to give it another push. Julian was always there to give me a nice kick in the butt when it was needed and, especially, she kept me sane through this process by reminding me that life must be also lived (free of guilt!), not only worked and researched. María Sierra is a soul mate who shared with me not only the experience of being a graduate student but also of moving abroad and building a new life while staying true to oneself.

I am also deeply grateful to my parents and sister, for engraving in me hard work ethics and the value of knowledge, for always being there – even from a distance – and, most importantly, for teaching me never to quit. Their strength, and the family stubbornness,

brought me here.

My Canadian family was also always a source of encouragement. I appreciate their passion for life and their dedication to each other. Thanks to – and forced by – proximity, *matante* Charlotte played a special role: her suppers and regular progress enquiries kept the research engine going.

Last but not least, I want to thank the person that was there every day and every night, who shared laughs, tears, glasses of wine, and late nights on the computer; the person who sacrificed countless evenings and weekends: my caring, loving, wittingly annoying and delightful husband, Stéphane.

Table of contents

1 Introduction.....	- 7 -
1.1 Background.....	- 9 -
1.1.1 A close-up look at translation environment tools (TEEnTs).....	- 9 -
1.1.2 From term banks through TMs to integrated termbases	- 18 -
1.2 Objectives.....	- 20 -
1.3 1.3 Hypothesis and sub-hypotheses.....	- 20 -
1.4 1.4 Methodology overview	- 22 -
1.5 Understanding this project’s limitations.....	- 23 -
1.5.1 Limitations on termbase use	- 24 -
1.5.2 Limitations on sample control.....	- 24 -
1.5.3 Limitations on target respondents	- 25 -
1.5.4 Limitations on hypothesis testing.....	- 26 -
1.5.5 Limitations on the resulting guidelines.....	- 27 -
1.6 1.6 Justification and motivation for research.....	- 29 -
1.7 1.7 Outline.....	- 31 -
2 Literature review	- 34 -
2.1 Terminology management: The what and why.....	- 34 -
2.1.1 What is terminology management?.....	- 34 -
2.1.2 Terminology management: a translator’s task?	- 35 -
2.1.3 Why manage terminology?	- 37 -
2.1.4 Obstacles to terminology management.....	- 41 -
2.2 Terminology management: the how	- 43 -
2.2.1 Terminology management in the translation literature.....	- 44 -
2.2.2 Terminology management in the translation technology literature.....	- 47 -
2.2.3 Terminology management in terminology literature.....	- 49 -
2.2.4 Terminology management in practice	- 59 -

3 Preliminary hypotheses..... - 61 -

3.1 Sub-hypothesis A: - 64 -
3.2 Sub-hypothesis B: - 67 -
3.3 Sub-hypothesis C: - 69 -
3.4 Sub-hypothesis D:..... - 70 -
3.5 Sub-hypothesis E: - 72 -
3.6 Sub-hypothesis F:..... - 76 -
3.7 Sub-hypothesis G:..... - 78 -

4 Methodology - 81 -

4.1 Selecting a data collection approach - 81 -
4.2 Selecting a type of survey..... - 84 -
4.3 Selecting a sampling design - 88 -
4.4 Analyzing survey data..... - 89 -

5 Terminology management practices among integrated termbase users - 93 -

5.1 Survey introduction: Use of terminology management systems integrated with translation environment tools - 93 -
5.2 Target respondents and criteria for participation - 96 -
5.3 Survey design - 97 -
5.3.1 Form - 97 -
5.3.2 Content..... - 98 -
5.4 Survey distribution..... - 100 -
5.5 Survey results - 101 -
5.5.1 Respondents' profile..... - 102 -
5.5.2 Terminology management planning - 125 -
5.5.3 Content selection - 138 -
5.5.4 Information-recording strategies..... - 144 -
5.6 Discussion - 152 -
5.6.1 Respondents' profile: professional..... - 153 -
5.6.2 Respondents' profile: technological..... - 156 -
5.6.3 Respondents' profile: training..... - 158 -
5.6.4 Respondents' profile: perception - 161 -

5.6.5 Terminology management planning: usage	- 162 -
5.6.6 Terminology management planning: Guidelines	- 164 -
5.6.7 Terminology management planning: Database organization.....	- 166 -
6 Evaluation of preliminary hypotheses	- 169 -
6.1 Sub-hypothesis A	- 169 -
6.2 Sub-hypothesis B.....	- 171 -
6.3 Sub-hypothesis C	- 172 -
6.4 Sub-hypothesis D.....	- 175 -
6.5 Sub-hypothesis E	- 177 -
6.6 Sub-hypothesis F.....	- 178 -
6.7 Sub-hypothesis G.....	- 179 -
6.8 Summary of sub-hypotheses to test further	- 180 -
7 User-acceptance test	- 181 -
7.1 Target audience and criteria for participation	- 181 -
7.2 Survey design.....	- 182 -
7.3 Survey distribution.....	- 184 -
7.4 Survey results.....	- 186 -
7.4.1 Respondents' profile.....	- 187 -
7.4.2 Recording synonyms	- 189 -
7.4.3 Recording non-term units	- 194 -
7.4.4 Recording multiple term forms	- 195 -
7.4.5 Using TBX-Basic as a term record template	- 196 -
7.4.6 Sub-group analysis	- 198 -
7.5 Discussion.....	- 200 -
7.5.1 Sample and respondents' profile	- 201 -
7.5.2 Recording synonyms	- 202 -
7.5.3 Recording non-term units	- 206 -
7.5.4 Recording multiple term forms	- 209 -
7.5.5 Using TBX-Basic as a term record template	- 214 -

8 Integrated-termbase design guidelines	- 218 -
8.1 Applying the guidelines	- 232 -
9 Concluding remarks	- 239 -
9.1 Assessment of the objectives	- 239 -
9.2 Lessons learned	- 240 -
9.3 Call to the community.....	- 242 -
9.4 Future research.....	- 244 -
9.5 Closing remarks.....	- 246 -
10 Bibliography	- 247 -
APPENDIX A: University of Ottawa Research Ethics Board Approval Notices-	263 -
APPENDIX B: Description of Existing Surveys on TM Systems Usage and Terminology Management	- 268 -
APPENDIX C: Use of Terminology Management Systems Integrated with Translation Environment Tools Survey Questionnaire	- 279 -
APPENDIX D: Integrated Termbases Optimization Survey Questionnaire.....	- 316 -
APPENDIX E: Use of Terminology Management Systems Integrated with Translation Environment Tools Survey Results	- 342 -
APPENDIX F: Integrated Termbases Optimization Survey Results.....	- 397 -

Figures

Figure 1 Example of a Hybrid Text Resulting from Terminology Pretranslation	- 17 -
Figure 2 Respondents' Areas of Specialization.....	- 104 -
Figure 3 Distribution of the Use of Different TEnT Tools.....	- 108 -
Figure 4 Distribution of Main TEnT Tools Used.....	- 110 -
Figure 5 Experience Using TEnTs.....	- 112 -
Figure 6 TMS Coverage by TEnT Training Provider.....	- 116 -
Figure 7 Content Recording Coverage by TEnT Training Provider.....	- 117 -
Figure 8 TMS Weight in TEnT Selection.....	- 120 -
Figure 9 Level of TMS Knowledge.....	- 121 -
Figure 10 Guideline Limitations on Recordable Units.....	- 132 -
Figure 11 Comparison of Unit Selection Criteria Ranked as Very Important.....	- 139 -
Figure 12 Unit Types Recorded.....	- 140 -
Figure 13 Reproduction of Durán Muñoz's table "What do you think a good terminological resource for translators should offer?"	- 149 -
Figure 14 Respondents' Age Distribution.....	- 156 -
Figure 15 TL Synonym Display (Full Form) during Interactive Translation.....	- 190 -
Figure 16 TL Synonym Display (Acronym) during Interactive Translation	- 191 -
Figure 17 Equivalent Insertion during Pretranslation in cases with Multiple TL Synonyms	- 192 -
Figure 18 Examples of Non-Term Units	- 194 -

Tables

Table 1 Example of Aligned Segments.....	- 11 -
Table 2 List of Previous Surveys on Terminology Management and TEnT use	- 60 -
Table 3 Breakdown of Survey Questions.....	- 97 -
Table 4 TEnT Usage as Reported in LISA (2004), Lagoudaki (2006) and This Study.....	- 113 -
Table 5 Presence of Guidelines in each Work Setting.....	- 129 -
Table 6 Use of Resources for Guideline Creation	- 130 -
Table 7 Guidelines Resources with Inconsistent Results.....	- 131 -
Table 8 Presence of Grammatical Fields in Respondents' Record Templates	- 145 -
Table 9 Record Template Model	- 146 -
Table 10 LISA 2005 Results on Recorded Information.....	- 147 -
Table 11 Participants per Country.....	- 188 -

1 Introduction

The principal goal of this research is to investigate and evaluate the strategies available for optimizing terminology management for translation purposes *within translation environment tools* (TEEnTs).

Before discussing the core of this research, it is useful to define the terms *terminology*, *terminology management*, *localization* and *translation* as they will be used within this project.

Terminology has several meanings. Terminology refers to “the set of special words belonging to a science, an art, an author, or a social entity” (Pavel and Nolet, 2001, p. xvii) – i.e. the set of terms used in a particular domain. At the same time, terminology is “the language discipline dedicated to the scientific study of the concepts and terms used in specialized languages” (Ibid.).

Terminology management refers to the activities necessary to generate and maintain collections of terminological information. Sager (1990, p. 2) defines such activities as the “collection, description, processing and presentation of terms, i.e. lexical items belonging to specialised areas of usage of one or more languages”. Some authors also use the term *terminography* to refer to terminology management. *Terminography* is a less popular denomination in English but its French equivalent, *terminographie*, has been widely accepted (L’Homme, 2004; pp. 15-16).

Localization, according to Esselink (2000, pp. 2-3), refers to the process of linguistically and culturally translating and adapting a product, including the software application and all supporting documentation, to a target locale (i.e. the target language and culture of a specific region).

Translation, in turn, refers to the transfer of material from one language to another (Esselink, 2000, p. 2).

The key differences between these two concepts are that *localization* involves the translation of electronic resources and its documentation while *translation* applies to any text, and *localization* involves a project with a number of activities and participants (e.g. multilingual project management, software and online help engineering and testing, translation strategy consulting) including the actual *translation* of texts (Ibid.).

We are well aware not only of the differences between *translation* and *localization*, as described above, but also of the fact that the translation process will be affected by a different set of constraints and needs depending on whether it takes place in a translation or a localization project. For example, when translating electronic resources, the target text may have a character length limit to fit in specific interface areas (e.g. buttons, menus, captions). Moreover, sharing decisions on source and target language terminology among programmers, marketers and translators becomes much more relevant given that the electronic resource, the marketing material, the supporting documentation and all target language versions of these materials are often developed in parallel. However, for the purpose of this research, the term *translation* will refer to any process by which text is transferred from one language to another, including localization.

Given that translation is an activity within localization, we believe that the resulting guidelines will still be useful for localization projects. Future research could include a more targeted study to evaluate whether the resulting guidelines are fully transferable to localization projects and if not, which modifications are necessary. However, for the current project, we will look into the subject from a more general point of view.

1.1 Background

A crucial aspect that sets this research apart from other work done in terminology management is its contextualization within TEnTs. Therefore, it is important to begin by providing a basic introduction to the terminology and functionality associated with TEnTs and their key components as well as to present a general idea of the market needs and professional situation that call for the implementation of these tools. Once we have outlined this essential background knowledge, we will provide a more detailed introduction to the specific objectives of this project and to the principal motivations for carrying out this particular research.

1.1.1 A close-up look at translation environment tools (TEnTs)

Translation environment tools, as the word “environment” in their name suggests, are software programs that provide an integrated framework for a variety of features and functions to support translators’ work. Generally, a TEnT will include at least a translation memory (TM) database and a terminology database or *termbase* (i.e. a collection of systematically organized term records), both of which are directly linked to a text editor. Additional tools may be present in a TEnT, and depending on the particular product in question, these could include automatic term extractors, text analyzers, spell-checkers, concordancers, machine translation systems, and so on. This collection and integration of a range of tools is based on the “one-stop shopping” principle, and the resulting product is generally known in the computer science world as a *tool suite*.

Although the initial idea of a TEnT dates back to the late 1970s and early 1980s (Arthern, 1979, p. 93; Kay, 1980 [1997]), such tools were only widely commercialized toward the late 1990s. These tools were brought to the market in response to the increasing demand

for translation that had come hand in hand with the advent of new technologies facilitating text publication and distribution, along with the far-reaching business trend of globalization¹.

Traditional translation approaches could not keep up with this boom, on one hand because of the magnitude of the increase in demand, and on the other hand, as a result of the fact that this surge in demand coincided with a shortage of professional translators (AILIA, 2004, p. 3; CTISC, 1999, p. 31). TEnTs allow users to store past translations and glossaries in order to recycle them if a similar text needs to be translated in the future. The underlying idea is to increase translators' productivity and consistency, which can in turn help them better meet the increasing demand for translation.

The key resource in a TEnT, around which the other elements of the collection are built, is without a doubt the *translation memory* (TM)². The TM consists of a “database of previous translations” (Somers, 2003, p. 31). In this database, legacy translations (i.e. previously translated texts) are paired with their corresponding source texts. Each pair of texts is broken down into small units (*segments*) – usually at the level of sentences or sentence-

¹ Note that the terms *globalization* and *internationalization* differ in meaning when used in the field of business management or translation and localization. For definitions of these concepts from the translation and localization industry perspective, refer to LISA's *Globalization Industry Primer* (2007, pp. 1, 11, 19), to Esselink (2000, pp. 2-4) or to Pym (2004, pp. 29-37). The above-cited works also provide in-depth descriptions of how globalization entered the translation industry and the new sectors it has created. For a more sociological overview of the impact of globalization on translation, refer to Cronin (2003). Readers can find definitions of these concepts from the business-management perspective in Parker's *Globalization and Business Practice* (1998, p. 51).

² Given that TMs are the cornerstone of TEnTs, a metonymic relation has developed whereby a TEnT is often known or referred to by the name of its principal component, a TM. In other words, some authors use the term TM to refer to an entire TEnT.

Due to a similar metonymic relation, readers may also find the acronym TM being used to refer to the TM system, the software application used to identify matches between a new source text and the contents of the translation memory database.

Because this research focuses on terminology management within TEnTs as a whole, and not simply within the TM component, we have opted to use the term TEnT to stress the tool suite nature of this type of application. The term “TEnT” has appeared in the translation technology literature, along with other competing terms, since the tool was first conceived. However, recently, TEnT has been very strongly championed by Jost Zetzsche (2006). Other terms for a TEnT that commonly appear in the literature are *translation workstation* (Melby *et al*, 1980), *translator's workstation* (Hutchins, 1998) or *translator's workbench* (popularized by Trados, one of the largest TEnT distributors that has since merged with SDL to form a new company known as SDL Trados).

like units – and each source segment is linked (*aligned*) with its translation in the target language (see Table 1).

Source Segment	Target Segment
All applications will be acknowledged.	Nous accusons réception de toutes les demandes d'admission.
* A reference number and personal identification number (PIN) will be assigned to each applicant. These numbers can be used to check the status of an application online at www.infoweb.uottawa.ca	* Un numéro de référence et un numéro d'identification personnel (NIP) sont attribués à toutes les candidatures. Ces numéros permettent de vérifier l'état de la demande d'admission à l'adresse www.infoweb.uOttawa.ca .
* Only applications with all required documents will be evaluated.	* Seuls les dossiers complets sont évalués.
* Incomplete applications will not be processed and may be cancelled.	* Les dossiers incomplets ne sont pas traités et peuvent être annulés.

Table 1 Example of Aligned Segments

Of course, having a collection of aligned texts is useful only if the contents can be easily searched. Therefore, along with the database, a TEnT includes a retrieval system that takes a new source text that a translator must work on and automatically compares it to those texts in the TM that have been previously translated as well as to the termbase in order to identify and retrieve any repeated passages (EAGLES, 1996). This system can search for the exact segment or a similar³ one in the TM. If there is any match, the translations of those segments or terms are automatically presented to the translator, who can assess whether and to what extent they can be reused in the new text to be translated.

³ Segments that are similar, but not identical, to those found in the TM are often referred to as “fuzzy matches”. More detailed information about different types of matches and on the general functioning of a TM can be found in references such as Bowker (2002a), Somers (2003) and L’Homme (2008).

1.1.1.1 TEnT translation match types and their display ranking

When translating, available results from both the TM and the termbases are displayed. However, the tool must select which match is more relevant to be inserted automatically in the translation pane as its top proposal. This choice is governed by a set of rules, which vary depending on each TEnT.

In MultiTrans, for example, if we need to translate the sentence “*Experience the power of market-leading engineering and cutting-edge design.*” the following logic will apply when ranking the available matches:

- 1. Exact term matches.** First the system will look in the termbase to check if the entire sentence exists as an exact term match. If the sentence were a slogan, for example, the user might have entered it in the termbase to make sure that it is always translated consistently in order to preserve the brand image. The termbase matches are prioritized above any other resource because the user has intentionally created that record, which means not only that the equivalent retrieved is highly reliable but also that the user considers it useful to make the information available for retrieval.
- 2. Confirmed exact segment matches.** If no exact term match is found, the system will turn to the TM database in order to check whether this exact sentence has ever been translated before and if so, it will look for any occurrence for which the user has previously verified and approved the alignment. If such a match exists, it will be proposed as the best available translation. If more than one confirmed exact segment is available, the system will rank the matches according to the TM prioritization established by the user and according to the user’s choice to view newer or older segments first.

3. **Unconfirmed exact segment matches.** If no exact term match or confirmed segment match is found, but unconfirmed matches are available, the system will display these as its best options and it will rank them according to the criteria explained in point 2.
4. **Fuzzy segment matches.** If no exact term, confirmed or unconfirmed exact segment matches are available, the system will propose fuzzy matches as the best available translations, i.e. previously translated segments that resemble the source segment with a maximum difference that the user can set to a specific threshold, also known as fuzzy factor. In MultiTrans the fuzzy factor is calculated based on the number of words that repeat between the sentence in the TM and the sentence in the document to translate (e.g. if 9 out of 10 words are exactly the same, we will obtain a 90% match). Note that other systems calculate the fuzzy factor based on the number of characters that differ. So that if 9 out of 10 words are the same and one word is different only because of a plural marker, the fuzzy match will most likely be higher than 90% in a TEnT that calculates fuzziness at the character level. For example, at this stage the tool would propose segments such as “*Experience the power of market-leading ~~engineering~~ **technology** and cutting-edge design.*”⁴ If there are multiple fuzzy matches, these are presented in order of decreasing percentage of similarity; if they have the same percentage of similarity, they are presented according to the criteria explained in point 2.

⁴ Text that appears in the “strikethrough” typographic presentation indicates that a word was present in the match stored in the TM but is not present in the new sentence to be translated. Meanwhile, text shown in boldface indicates that a word is present in the sentence to be translated but does not appear in the match stored in the TM database.

5. **Partial term matches.** If none of the above match types are available, then the most reliable match information is considered to be term matches, be they single-word terms, multi-word terms or even phraseological expressions. For example, the user may have created records for *technology* and *design* but also for *cutting-edge design* or *market-leading technology*. This may be more useful to the user than sections of the segment that may be found in previously translated material such as *Experience the power* or *the power of market*.
6. **Sub-segment matches.** The lowest ranking type of match is the sub-segment match, i.e. a section of the segment present in previously translated documents that does not reach the minimum fuzzy match threshold. Such matches can be very useful if they turn out to be specialized terms or phrases difficult to translate (*market-leading technology*), but they can also be just a series of words that tend to occur together (*the power of market*). All available sub-segments for a sentence are displayed.

This description illustrates the ranking order of match types. If multiple match types exist, they will all be displayed (i.e. users will see that the system may have a term match, a confirmed match, an unconfirmed match, a fuzzy match and a sub-segment match available); however, results will be ranked as described above.

1.1.1.2 Terminology management systems in general and within TEnTs

The TEnT resource that will be the focus of this research is the *terminology management system* (TMS). The International Organization for Standardization (ISO) has a technical committee (TC 37) for “Terminology and other language and content resources”. This committee is currently evaluating the 2009 draft for what will be, if approved, the 2009 ISO/DIS 26162 draft entitled *Systems to manage terminology, knowledge and content – Design,*

implementation and maintenance of terminology management systems. In its current state, this draft defines a TMS as “a software tool specifically designed to collect, maintain, and access terminological data for use by translators, terminologists, and various other users” (p. 8). A TMS stores terminological entries within one or more termbases. In turn, termbases are files based on the principle of a database: instead of requiring the user to design the database structure from scratch, they come with predefined fixed, modifiable or fully customizable terminology record structures (L’Homme, 2008, p. 134; Bowker, 2002a, p. 78). ISO 1087 *Terminology Work – Vocabulary – Part 2: Computer Applications* defines a termbase as a “database comprising a terminological resource” and a terminological resource as a “text or data resource consisting of terminological entries” (ISO 1087-2:2000, 2.22).

Before proceeding with the description of integrated terminology databases in TEnTs, it is important to clarify the difference between the terms *termbase* and *term bank*, which are sometimes used interchangeably. The 2009 ISO/DIS 26162 draft does not define *term bank* as a concept, but in its introduction it describes a term bank as the largest type of terminology database, usually created by “major companies and governmental agencies” (2009, p. vii). A term bank will thus be created to reach a wide and heterogeneous audience that could include company’s staff, an association membership, or even the general public.

According to the above definition, a term bank is a type of terminology database; however, in the context of this research project, these terms will be considered to refer to two quite distinct concepts. The term *term bank* will be reserved exclusively for the larger databases to which it most typically refers (e.g. *TERMIUM*[®], *Le grand dictionnaire terminologique* (GDT), *InterActive Terminology for Europe* (LATE), the *United Nations’ Multilingual Terminology*

*Database (UNTerm))*⁵. The term *termbase* will be used to refer to an electronic collection of structured term entries in the form of individual or client-server databases of a relatively smaller size and with a more limited audience than a term bank. This is the term that best corresponds to the focus of this research: terminology databases created primarily within a TEnT by translators and for translators (either working independently or in a team) for the purpose of translation within a TEnT.

Another point that must be noted is that, as described in the ISO/DIS 26162:2009 draft (p. 9), TMSs can be *stand-alone*, *integrated* or *combined*. Stand-alone TMSs are fully independent tools whose sole function is the collection, storage, organization and retrieval of terminological information. Integrated TMSs are a component of a TEnT. In this context, TMSs not only serve the same purpose as a stand-alone system but also operate in conjunction with a TM to become the core resources that are actively searched during an interactive translation session with the tool suite. Combined TMSs can work as stand-alone tools or integrated with a TEnT. For the sake of clarity, when referring to a TMS within a TEnT and its database(s), I will use the terms *integrated TMS* and *integrated termbase*, given that the research questions that make up this project all revolve around TMSs that form part of a TEnT.

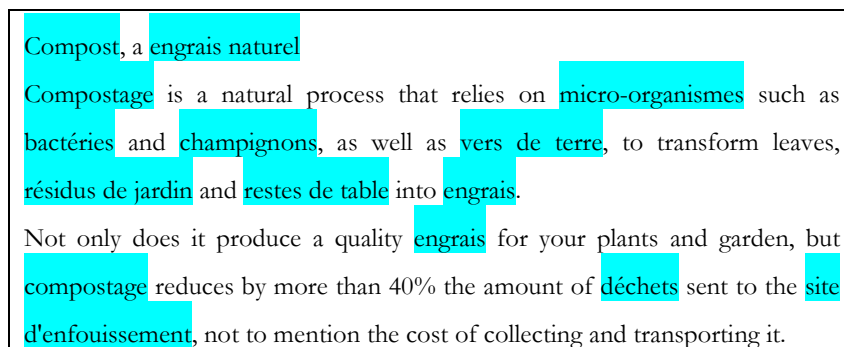
By storing data electronically in a structured way, all TMSs offer the user quick, easy and focused access to the information. Therefore, logically, TMSs also offer search mechanisms that allow the user to find targeted records using search queries that often can apply to the main entry alone, or to any other field in the record. Such search mechanisms usually are also compatible with wild cards (e.g. using an asterisk for word truncation or a question mark for a single character substitution), Boolean operators (e.g. using AND to

⁵ These term banks can be consulted online: TERMIUM® at <http://www.btb.termiumplus.gc.ca/>, the GDT at <http://www.granddictionnaire.com/>, IATE at <http://iate.europa.eu> and UNTerm at <http://unterm.un.org/>.

combine several conditions that resulting records must satisfy and OR to extract records that satisfy any of the conditions established) and fuzzy search algorithms (searching results that resemble the term entered but that are not exactly the same (e.g. color – colour, mode – modal, college – collage)) (Bowker, 2002a, p. 79; L’Homme, 2008, p. 136).

As mentioned above, integrated TMSs and TMs work as TEnT resources that may be combined with a series of extended features that aim to help the translator during the translation process:

- Active terminology recognition (Bowker, 2002a, p. 81; L’Homme, 2008, pp. 183, 185): This retrieval option consists in the TEnT proposing term matches for any expressions in the text to be translated that exist in the termbase. (This option can be used in conjunction with automatic retrieval of information from the TM.)
- Pretranslation (Bowker, 2002a, p. 81-82): This function scans a text to identify expressions that exist in the termbase or TM and replaces them with the recorded equivalents in batch mode. As illustrated in Figure 1, the result is a hybrid text in which certain terms and expressions are translated automatically, and the translator can then translate the remainder of the surrounding text and revise the complete document afterwards.



Compost, a engrais naturel
Compostage is a natural process that relies on micro-organismes such as
bactéries and champignons, as well as vers de terre, to transform leaves,
résidus de jardin and restes de table into engrais.
Not only does it produce a quality engrais for your plants and garden, but
compostage reduces by more than 40% the amount of déchets sent to the site
d'enfouissement, not to mention the cost of collecting and transporting it.

Figure 1 Example of a Hybrid Text Resulting from Terminology Pretranslation

- Term record creation within the translation workflow (Zetzsche, 2006): This function allows users to automatically create term records from expressions found in the TM (for example, by selecting the source and target term and clicking on a command to add those expressions to the terminology database) during the interactive translation of a text.

Integrated termbases differ from term banks created exclusively as general reference tools in which terms are manually looked up. Integrated termbases are used as reference resources for manual, semi-automatic and automatic information retrieval during the translation process⁶. Therefore, some basic facts may be very different:

- the database can be shared or personal;
- translators are the main contributors to such termbases (although in some cases the work is verified by a terminologist);
- the main contributors to these databases are generally also the main users;
- terminological information can be automatically stored and/or retrieved during the translation process.

1.1.2 From term banks through TMs to integrated termbases

Traditional term banks were among the first computer-aided translation tools to be developed, dating back to the 1960s. As such, they have been documented in great detail, including discussions of their design, contents, applications, benefits and drawbacks (e.g.

⁶ Not all terminology databases created within a TEnT are intended to be exploited interactively for translation. The terminology management system of these tools can be used to create terminology databases that will serve as a reference resource independent of any other tool of the TEnT. In that case, those databases are closer to stand-alone general reference terminology databases. Examples of terminology databases integrated to TEnTs used to manage a field-based or general traditional term bank are TERMDAT, the term bank of the Swiss Federal Administration, and Universal Postal Union official termbase TERMPOST, both powered by the terminology component of *MultiTrans*, MultiCorpora's TEnT (<http://www.multicorpora.com>).

Rondeau, 1984; Sager, 1990; Pavel and Nolet, 2001).

In the next big wave of computer-aided translation development, TEnTs became widely commercially available in the late 1990s and, in less than a decade, a plethora of such products have flooded the market⁷. The TEnT software boom arrived, logically, hand-in-hand with a flurry of literature discussing the concept, reviewing the tools and evaluating their performance. Scholars have mainly focused on what TEnTs can do for the translator and the translation process: analyzing their capabilities, advantages, disadvantages, reception by the community, etc.

To date, TMs have been the star component of TEnTs and extensive research has been carried out on the best practices for managing these repositories of text. Based on these previous studies, translation scholars and practitioners have come up with a series of valuable recommendations to make working with TMs easier and more productive (e.g. Heyn, 1998; Webb, 1998; Benis, 1999; Trujillo, 1999; Wilss, 1999; Esselink, 2000; Arrouart and Bédard, 2001; Austerhöhl, 2001; Lanctôt, 2001; Bowker, 2002a, 2005; Bowker and Pearson, 2002; Okunev, 2005; Lagoudaki, 2006; Quah, 2006; Sofer, 2006; García, 2010; Pym, 2011).

To the best of my knowledge, however, no similar type of detailed research has been carried out with regard to the best use of an integrated termbase. Therefore, users do not have access to an established methodology or proven guidelines to help them to effectively manage and optimize termbases that are intended to serve as a linguistic resource within a TEnT. This research focuses on the terminology management and database component of TEnTs, in the hopes of taking a step towards filling this gap.

⁷ Examples include SDL Trados, Déjà Vu and MultiTrans. See Lagoudaki (2006, p. 18) for a list of the 36 most popular TEnTs.

1.2 Objectives

As noted above, the overall objectives of this research are to investigate and evaluate the current use of integrated TMSs and to suggest strategies for optimizing translation-oriented terminology management within these systems.

To achieve these overall objectives, a number of more specific goals are necessary:

- a) Identify current practices in the use of integrated TMSs among translators, and in so doing, find out whether translators use this tool within the TEnT, and if so, how they design their integrated termbases, what sources they use as references on how to design and build their termbases and, finally, how they put their termbases to use.
- b) Identify user-controlled strategies currently being applied by users or strategies that can be developed based on existing literature and personal experience and that can contribute to optimizing terminology management within TEnTs in order to obtain better results during translation (i.e. a higher number of relevant terminology matches).
- c) Evaluate user acceptance of the strategies identified in point b).

1.3 1.3 Hypothesis and sub-hypotheses

The above-mentioned objectives are motivated by an underlying hypothesis. This hypothesis stems from anecdotal observations on how translators manage terminology within TEnTs. Our observations seemed to indicate that:

Translators working interactively or pretranslating with TEnT termbases adapt the terminographical method to their specific needs: automated term lookup and one-click insertion in the translated text.

As we started evaluating the possibilities of this hypothesis, we came across specialized

articles that pointed in the same direction (Kenny, 1999; O'Brien, 2001), which encouraged us to research this matter further.

After carrying out the literature review and further reflecting on the issue at stake, we formulated a series of sub-hypotheses on how the general hypothesis could materialize:

- a) Contrary to what current terminology and terminography literature recommends, translators will use fewer term record fields in a TEnT-integrated termbase.
- b) Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard.
- c) Contrary to what current terminology and terminography literature recommends, translators will classify records in personal TEnT-integrated termbases first by client or project and only secondly by domain.
- d) Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.
- e) Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.
- f) Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will not be opposed to extracting terms/units and equivalents from translated texts.
- g) Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or in their most frequent form(s).

This hypothesis and its sub-hypotheses will be presented in detail in chapter 3. The approach followed to research and test them is described in section 1.4, and further details can be found in chapter 4.

1.4 1.4 Methodology overview

In order to achieve the objectives presented in section 1.2, this research followed the path described below.

First, before venturing into uncharted territory, one must be familiar with the ground covered so far. The project started with a review of the existing literature on the nature of terminology management and its relevance within the translation process. Once the basic concepts and the role of such practice were established, it was important to collect any information available on terminology management for translators working with a TEnT, the main focus of this research. Given the scarcity of such literature, we turned to key works on traditional terminology management for terminologists, the closest body of literature available for the needs of this project.

Second, based on the literature review, personal experience and impressions, we formulated a series of preliminary hypotheses. These hypotheses set out terminology management practices that we predicted would contribute to the creation of integrated termbases that would be more useful for translators working with TEnTs by producing more satisfactory results than termbases built following traditional terminology management principles.

Third, in order to situate the preliminary hypotheses in the context of current usage of integrated termbases, we conducted a survey on TEnT usage, perception of terminology

management within TEnTs and approaches to terminology management. (See Appendix C for a copy of the full survey.) While this survey was built on existing surveys that have targeted TEnT and terminology usage (Lommel, 2002, 2004, 2005; eColoTrain, 2006; Lagoudaki, 2006; Kelly and DePalma, 2009) and further research on this topic (O'Brien, 1998; Kenny, 1999; Bowker and Marshman, 2009; Bowker, 2011), it differed from previous surveys and research by specifically exploring the design and use of termbases within TEnTs by translators, for translation purposes. The results of the above-described survey and corpus of existing surveys served to establish a portrait of the current usage of these tools by translators and the challenges they face.

Fourth, we analyzed the preliminary hypotheses in the light of the results obtained in the survey of current usage of integrated termbases in order to confirm, refute and/or refine the preliminary hypotheses, as well as to select which of them required further investigation.

Fifth, we carried out a second survey among users to test their acceptance of the preliminary hypotheses that could not be confirmed or refuted based on the analysis of current practices.

Finally, we formulated a series of proposed best practices that, based on the functioning and purpose of a TEnT and the actual use translators make of integrated termbases, should assist a translator in designing and using an integrated termbase to his or her best advantage.

1.5 Understanding this project's limitations

Given that this research was carried out within the framework of a doctoral thesis, certain limitations applied. These were mainly owing to time constraints and the human and financial resources available.

1.5.1 Limitations on termbase use

As presented in section 1.1.1.2, termbases can be stand-alone, integrated or combined. Integrated termbases occur within TEnTs. Within such environments, these termbases can be queried automatically by the system during interactive translation and pretranslation or queried manually as if they were standalone termbases.

This project focuses almost exclusively on how to optimize terminology management within TEnTs for translation in interactive or pretranslation mode only, during which terminology is queried automatically. This decision is further justified by the findings of the initial survey on terminology management practices within TEnTs, which revealed that only 15.4% of respondents used their TEnT as a reference tool to carry out manual searches (see section 5.5.1.2 for more details).

The effects of the optimization strategies on termbases with a double use as stand-alone reference tools, in which terms are manually queried, will not be included in this project. However, we acknowledge that such research would be not only interesting but also valuable.

1.5.2 Limitations on sample control

A number of different options are possible for gathering information about TEnT usage, such as creating focus groups, interviewing tool users, or undertaking case studies. However, in the context of this project, as explained above, we opted for a survey as the most practical means for gathering the greatest range of useful data.

As explained in section 4.3, we used a purposive no probability sampling design. Because the survey was announced by sending invitations by email and to discussion forums

and distribution lists, it is not possible to know how many people received the invitation. Therefore, it is likewise impossible to calculate a response rate. To mitigate this situation, and in order to give a clear picture of the number and type of participants who answered each survey, we included a respondents' profile section in each survey as well as a series of mandatory questions filtering out participants who did not meet the basic participation requirements.

1.5.3 Limitations on target respondents

Given that the ultimate goal of this project is to formulate best practices for the creation and use of translation-oriented integrated termbases within a TEnT, both surveys aimed to reach TEnT users.

In order to ensure we obtained a usable sample for the surveys, each survey started with a series of mandatory filtering options. A usable sample in this case required not only that the participants be TEnT users but also that those participants be capable of answering the English-language survey and be willing for their contributions to be used for academic research.

Therefore the filtering criteria were as follows:

- Consent to participate in the survey. To meet this criterion, in the first survey, participants were asked to accept that the results of the survey be used for research and future publication purposes, and in the second survey participants were asked to read and accept a consent form detailing all the specifics about the survey (goals, risks, period during which the data will be

kept, data destruction method, etc.)⁸.

- TEnT usage. In both surveys, participants were asked whether they used a TEnT. Not using a TEnT led to the automatic exclusion of the participant from the survey.
- English reading comprehension. Owing to the limited resources available for this research, the surveys were distributed in English only. English was chosen as the lingua franca on the basis that the translator community is by nature polyglot and the fact that English seemed likely to be a language that a large number of people would be able to understand as a native or foreign language. However, this required all participants to have good reading comprehension in English.
- Adulthood. This limitation was adopted, as encouraged by the Research Ethics Board, in order to ensure that respondents were able to give consent to their participation in the survey.

1.5.4 Limitations on hypothesis testing

Given the limited time and resources available, hypotheses that were developed based on the literature review and/or results of the initial survey were not tested for their empirical impact on productivity, efficiency or quality. The preliminary hypotheses were first compared to the attested usage of these tools collected as part of the survey in step three. If any of the proposed hypotheses proved to be in wide use among the community (see section 4.4), the hypothesis was considered to be validated. The assumption behind this was that

⁸ The difference in the approach used to seek the respondents' acceptance to participate in surveys is owing to a change in the regulations set out by the University of Ottawa's Research Ethics Board.

only those practices considered to successfully provide acceptable results would be widely adopted by users.

The preliminary hypotheses that could not be validated or rejected based on the comparison with current usage trends were tested in a second survey as part of step five of this research. The second survey sought to determine whether TEnT users agreed or disagreed on the value that this set of strategies can add to an integrated termbase. These results do not empirically prove nor disprove the validity of these strategies (i.e. we did not obtain statistics on their impact on productivity, efficiency or quality). However, at the end of this second stage, we arrived at a series of proposals for best practices that in principle will contribute to the optimization of integrated termbase design and usage and whose usefulness has been preliminarily assessed thanks to the judgement of a sample of knowledgeable and experienced TEnT users.

In conclusion, hypotheses have been tested only conjecturally within the framework of this research. Empirical tests would certainly be a natural progression for this line of research. However, given the obstacles for this type of testing presented previously, such a test could target only one specific user sub-group at a time. As the aim of this research was to obtain broader guidelines, we sacrificed quantitative data on the impacts of the guidelines in practice for a wide scope of application.

1.5.5 Limitations on the resulting guidelines

All of the above limitations had an impact on the resulting guidelines. Moreover, language and translation have long proved to be very complex objects of study. Therefore, translators must always use their common sense to evaluate each guideline against their own

needs and decide whether it applies to their specific language pair, subject specialization, tools used and client needs.

Special attention must be brought to the fact that we used a nonprobability sampling design for our surveys (that is, respondents were not selected from the total population of potential respondents according to a specifically designed random selection process). This means that we cannot calculate response rates, the range of error or the level of confidence that would establish that the sample is a true representation of the population (Aday, 1996, p. 116; Manfreda et al., 2011, p. 984; Schonlau et al., 2002, p. 106). Therefore, we cannot infer generalizations about the population based on the data of our survey (AAPOR, 2011, p. 38; Bethlehem & Biffignandi, 2012, p. 445).

However, nonprobability sampling-based surveys can be used, as in this research, to study groups of users whose population frame (that is, information about total population of potential respondents) is hard to establish, and to draw hypotheses that will need to be tested at a future point using a probability-based survey or testing method (Aday, 1996, p. 116). This methodological constraint is further developed in section 4.4.

Nonprobability based surveys sufficed for the purposes of this research as the main goal was to propose guidelines for best practices that, as indicated in section 0, would be based on existing literature and personal experience, and that would be tested only conjecturally. Empirical tests on efficiency and productivity would be the logical next step. Unfortunately, owing to constraints on time and resources, such tests were beyond the scope of this project.

Moreover, it must be noted that the user acceptance test of the preliminary hypotheses in step five asked participants to answer the survey by placing themselves in a generic scenario where they would be using a nameless TEnT in a personal context, i.e. using

an individual integrated termbase and not a shared one. This limitation was added to avoid ambiguous responses that would depend on what working scenario participants had in mind. We opted for the non-shared integrated termbase scenario both because it reduces the complexity of the scenario by eliminating factors primarily related to exchanging termbases and because, based on the profile of respondents who completed the first survey, it was expected that at least half of the participants in the second survey would be freelancers, who were expected to be very familiar with these individual termbases, but less aware of the challenges of shared ones. It was assumed that it would be easier for users working with shared termbases to put themselves in a situation where the termbase would not be shared, rather than the reverse.

1.6 1.6 Justification and motivation for research

In recent years, there has been a growing awareness of the importance of terminology management within the translation profession. Several studies have shown that investing time and money in terminology management makes good business sense and provides a good return on investment (Champagne, 2004b; Childress, 2007; Wittner, 2007; Kelly and DePalma, 2009). In this context, it is increasingly being recognized that terminology has an important role to play, among other aspects, in

- enhancing the quality of source and target documents and increasing productivity in terms of facilitating the creation of accurate documentation and its translation, and ultimately helping all users of the documentation to better understand the message and ultimately to better understand each other when communicating on the subject matter,

- ensuring better communication of a company's brand, thus facilitating the interaction of different branches of a company (product design, product development, marketing, sales, customer service) and their interaction with their current and prospective customer base, and
- avoiding potential legal liabilities resulting from imprecise, incorrect or contradictory interpretation of documentation.

For these reasons, a project that seeks to investigate and optimize terminology management strategies is expected to be welcomed in the translation industry⁹. Specifically, as noted above, carrying out this research project will be beneficial for the field of translation studies because it fills a gap in our knowledge about TEnTs and their usage. A comprehensive and systematic research project on the nature and relevance of terminology management, on the current approaches applied to manage terminology within TEnTs, as well as an exploration of possible strategies to optimize that usage would expand our understanding of integrated termbases and the strategies available to optimize their use. In particular, this research will provide

- a) an overview of the **perception** that translators have of integrated termbases and the **use** they make of them as part of their translation process, and
- b) a set of **best practices** that can be applied to **optimize** these termbases.

The findings resulting from this research will simultaneously benefit translators, translator trainers and trainees, and translation software developers. Translators often acquire TEnTs attracted by the buzz in the community that praises increases in productivity and consistency. However, these tools come with little advice on how to build their required

⁹ In spite of its value, terminology management is viewed with a certain reticence, as will be described in the section 2.1.4 *Obstacles to terminology management*.

databases (usually a TM and a termbase). Unfortunately, little information is available in the academic world regarding integrated termbases. The results of this research will help translators not only to make a better-informed decision regarding the weight of the termbase within the TEnT but also to design their termbase from day one in order to benefit from it more fully. Such descriptions and guidelines will also provide translator trainers with another resource to educate translation students in the difference between stand-alone and integrated termbases as well as to guide them in making the best possible use of each in order to fully master the tools available to them in the current market. Traditional general term banks are quite well documented, but more information and guidance is needed to help trainee translators to build and use integrated termbases effectively. Finally, although the best practices proposed by this research are aimed at the end-user, software developers can view them as an inventory of areas of their tools that need improvement. If end-user practices can contribute to optimizing results within the TEnTs, software developers will certainly be able to adapt their tools to facilitate such practices. In the best of cases, they may even be able to eliminate the obstacles that the best practices attempt to work around.

The relevance of terminology management within the translation process will be discussed in detail throughout the literature review in chapter 2.

1.7 1.7 Outline

As presented in section 1.4, the key elements of this thesis are the review of the literature, the drafting of the hypotheses, the design and implementation of the two surveys used to test the hypotheses, and the production of the guidelines. The rest of this document will present the findings for each of these steps of the process. Accordingly, this thesis is

divided into nine chapters.

Chapter 1: Presented an overview of the background, the main objectives, hypothesis, methodology, limitations and motivation.

Chapter 2: Focuses on the literature review. Readers will find a summary introduction to the concept of terminology management (a description of what it is, who carries out terminology management, why they do it or why they do not), followed by a summary of terminology management practices described in translation, translation technology and terminology literature. Finally, the chapter enumerates a series of recent surveys on terminology management and translation memory usage, which are briefly described in Appendix B.

Chapter 3: Introduces the overarching hypothesis of this thesis as well as each of the sub-hypotheses (A to G), contextualizing each of them with concrete examples.

Chapter 4: Describes the methodology applied to this thesis research in greater detail. More precisely, it explains the decision process about which data collection approach and which survey type were selected as well as the approach followed to analyze the data.

Chapter 5: Presents the design of the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* on current practices to manage terminology within TEnTs, reports on the results obtained and discusses on the significance of its findings.

Chapter 6: Evaluates the preliminary hypothesis and sub-hypotheses against the results of the *Use of Terminology Management Systems Integrated with Translation*

Environment Tools Survey in order to determine if any sub-hypotheses can be confirmed or rejected based on current practices.

Chapter 7: Presents the design of, reports on and discusses the results of the *Integrated Termbases Optimization Survey*, which aimed at testing the user acceptance of the sub-hypotheses that could not be confirmed or rejected based on the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* results.

Chapter 8: Proposes a set of integrated termbase design guidelines to optimize the use of this type of linguistic resource within a TEnT and for translation purposes.

Chapter 9: Concludes this thesis with a reflection on the unfolding of this project, followed by a look towards future research that could be undertaken.

2 Literature review

This chapter will introduce terminology management as a concept and will provide an overview of who carries out terminology management, why language professionals should be interested in this practice and what we know of terminology management in the fields of translation studies and terminology as well as of terminology management carried out in the language industry.

2.1 Terminology management: The what and why

This section will first introduce the concept of terminology management as it is understood for this project and then discuss the role terminology management plays in the translation process and the reasons why it is important for translators and translation services to manage their terminological information in order to ensure language quality and accuracy, facilitate translation and reduce multiple searches and correction costs further down the document production chain.

2.1.1 What is terminology management?

As defined in section 1, terminology management groups all the tasks involved in designing and collecting data to create and maintain a terminology resource.

Terminology work can be carried out with various purposes and in a wide range of contexts. Firstly, it can be descriptive, when it is intended to “document all terms used to designate the concepts treated in a single discipline” (Wright, 1997, p. 18). In contrast, it can be prescriptive if it aims to standardize the terms used to denote concepts within a discipline

(Cabr , 1992, p. 33) or to establish and promote correct terminology for a language as part of language planning initiatives (Ibid., p. 35). Secondly, terminology work can be carried out systematically (thematically) or in an ad hoc fashion. Thematic terminology projects aim to document the terminology of a subject area relatively comprehensively. This is the approach to terminology work that is most widely known and practised within the discipline of terminology proper, inspired by exhaustive theoretical reflection such as the General Theory of Terminology (GTT) developed by Eugen W ster and expanded by numerous terminologists after him. This is also the terminological approach that results in the creation of specialized dictionaries and term banks.

In contrast, ad hoc terminology work is typically carried out either by translators or by terminologists working for a translation service or a company's linguistic service. These translators and terminologists need to answer isolated source- or target-language terminological questions (Dubuc, 2002, p. 41) that occur in texts coming from a range of subject fields (Wright, 1997, p. 19).

Integrated termbases will consist mostly of a series of ad hoc terminological records. These termbases generally store terms that posed a challenge during a given translation project (e.g. comprehension or translation difficulty) or presented a high frequency of occurrence within a text.

2.1.2 Terminology management: a translator's task?

The relevance of terminology work for translation is undeniable. All texts, and in particular specialized texts, contain terminology that translators will need to render in the target language. Hence, translators are often (and with good reason) identified as important

users of any terminological work produced by terminologists, such as term banks and glossaries. However, recently there has been a growing interest in terminology management as carried out by translators and/or within translation services.

One example of such interest is the April/May 2007 issue of *MultiLingual* – a professional publication that specializes in language, technology and business – which was dedicated to the topic of terminology management. Others include the Terminology Special Interest Group that was created by the former Localization Industry Standards Association (LISA) with the aim of promoting “terminology management as an essential part of the content development, globalization, internationalization, localization and translation processes” (LISA, 2008), as well as the surveys on terminology (Ibid., 2001) and terminology management (Lommel, 2005) that were been carried out by this association. Yet more examples are provided by the studies stemming from the largest translation service in Canada, the federal government’s Translation Bureau, on the role of terminology in Canada (Champagne, 2004a) and on the economic value of terminological work (Champagne, 2004b).

Terminology work is part of the translation process. The actual percentage of their working time that translators claim to dedicate to these tasks varies from survey to survey. According to LISA’s survey on terminology management, 50% of the respondents spent less than 200 hours per year (i.e. less than 10% of their working time) on terminology tasks (Lommel, 2005, p. 2); however, the study on *The Economic Value of Terminology* concluded that experienced translators invest 20-25% of their working time in terminology tasks, while inexperienced translators can invest up to 40-60% (Champagne, 2004b, p. 30). The fact remains that many translators do spend a portion of their time carrying out terminological tasks as part of the translation process. Champagne’s (2004a p. 26) study on the status of

terminology in Canada concluded that terminology tasks are a part of all stages of the “language process” (before, during and after writing or translating a document), and the descriptions provided by Jaekel (2000, pp. 159-179) and Joscelyne (2000, pp. 81-95) of the roles of the terminology and translation departments, within Ericsson and the Organization for Economic Cooperation and Development respectively, support this assertion.

In any case, even when companies have a terminology service – 7.8% of Canadian businesses do, although the rates increase with the size of the business (Champagne, 2004a, p. 17) – only between 0.3 and 1% of businesses employ an in-house terminologist (Champagne, 2004a, p. 19). That is to say, even when companies have a separate terminology service, a majority of the time language or product specialists other than terminologists carry out the terminological work. More precisely, 15% of Canadian small and medium enterprises (SMEs) have a terminology specialist in-house and in 16% of these cases it is a translator, which is equal to the percent of cases in which this position is occupied by a terminologist. In contrast, 19% of large Canadian companies have a terminology specialist in-house and in 29% of this cases it is a translator – while only in 20% of the cases is the specialist a terminologist (Champagne, 2004a, p. 18).

In short, translators not only carry out terminology tasks as part of the translation process in general, especially in companies where there is no terminology service *per se*, but they also often play the role of terminology specialists when companies do have a terminology service or when they work independently as freelancers.

2.1.3 Why manage terminology?

The reasons that have traditionally motivated terminology work in general are the

enhancement of language quality, accuracy and consistency. The quality-assurance role of terminology work is widely recognized by firms and within the language service industry, as the surveys carried out in recent years demonstrate (Champagne, 2004a, p. 26; Champagne, 2004b, p. 8; Lommel, 2005, p. 3; Dunne, 2007, p. 33; Kelly and DePalma, 2009, p. 8). This section will explore the factors that justify companies' and translators' attention to terminology and terminology management.

2.1.3.1 From a company's point of view

Managing a company's terminology has a series of benefits. Ensuring that translators and others keep an organized and easily retrievable record of term research and choices made ensures that the same term will be translated in the same manner whenever it appears in a similar context. Hence, terminology management promotes terminology consistency (Champagne, 2004a, p. 26; Champagne, 2004b, p. 12; L'Homme, 2008, p. 138; Kelly and DePalma, 2009, p. 8; SDL Trados, 2009a, p. 2).

Maintaining terminology consistency in source and target documents has multiple positive effects.

- a) It helps to promote correct use of terms (Champagne, 2004a, p. 26).
- b) It improves the quality of the final text (Champagne, 2004a, p. 31; Kelly and DePalma, 2009, p. 8; SDL Trados, 2009a, p. 5).
- c) It reduces time and effort invested in corrections (Champagne, 2004b, p. 31; Dunne, 2007, p. 37; Lommel, 2005, p. 3; Kelly and DePalma, 2009, p. 10), as the translator or reviser is not forced to search for and evaluate all potential equivalents used. This becomes increasingly relevant when working in multilingual projects as corrections and variations can grow exponentially when multiplied by a number of languages.

- d) It strengthens a company's brand and credibility (Champagne, 2004a, p. 26; Dunne, 2007, p. 37; Fidura, 2007, p. 41; Kelly and DePalma, 2009, p. 28; SDL Trados, 2009a, p. 2).
- e) It facilitates quality control of both products and processes (Champagne, 2004b, p. 12; Dunne, 2007, p. 37; Fidura, 2007, p. 41; Lommel, 2005, p. 3).
- f) It paves the way for clearer communication both within the company and with customers (Champagne, 2004a, p. 26).
- g) It contributes to reducing customer service calls and enhancing common understanding during those calls (Champagne, 2004b, p. 12; Dunne, 2007, p. 37; SDL Trados, 2009a, p. 2).
- h) Last but not least, it reduces the risk of product failure due to incorrect, ambiguous or inconsistent terminology and all liabilities attached to such undesirable events (Champagne, 2004b, p. 12; Dunne, 2007, p. 37).

At the same time, not managing terminology can have side-effects that are the exact opposites of its benefits.

- a) It promotes useless repetition of searches (Dunne, 2007, p. 33; Childress, 2007, p. 44; Lommel, 2005, p. 9), which translate into lower productivity.
- b) It increases the risk of terminological inconsistency. Firstly, if a translator or team of translators have used different term variants, then it can be a serious challenge to deliver a cohesive and harmonized text at the time of revision (Dunne, 2007, p. 33; Childress, 2007, p. 44). Secondly, it can lead to a considerable loss in efficiency resulting from a loss of product usability (Dunne, 2007, p. 33), which will damage the company's image and will hinder communications within the company and with

the client. This is likely to prevent the company from providing effective customer service, while at the same time creating confusion on the client's end which will increase customer service queries (Dunne, 2007, p. 33; Childress, 2007, p. 44).

2.1.3.2 From a translator's point of view

The same benefits that managing terminology brings to companies also apply in the case of translators. The difference is that translators will give more weight to certain benefits of terminology management.

The improved consistency that results from properly managing terminology is a key benefit for translators mainly because it increases the quality of the final text (Champagne, 2004a, p. 31; Kelly and DePalma, 2009, p. 8; SDL Trados, 2009a, p. 5). Moreover, a translator does not only need to ensure that accurate terminology is used; clients often request that a translator use their preferred equivalents or their proprietary terminology. Therefore, being able to systematically manage terminology in general and for individual clients is a must. A translator's reputation lies in the quality of the texts produced and his/her ability to respond to clients' requests; succeeding on both these fronts helps to ensure that clients request a translator's services again or that a translator is kept on staff. Terminological consistency also facilitates corrections (e.g. replacing one term with another) as it reduces the time it will take to locate all occurrences of a term in a text (Champagne, 2004b, p. 31; Dunne, 2007, p. 37; Lommel, 2005, p. 3; Kelly and DePalma, 2009, p. 10). Corrections may be requested by the client or may be implemented by the translator during translation or revision. If during the translation process multiple terms or variants have been used to translate a concept, implementing these corrections will be much more time-

consuming and more expensive. This is an expense that is rarely¹⁰ charged to the client, as translations often are quoted at a fee per word¹¹. The other key benefit of terminology management that affects translators is the elimination of useless repetition of searches (Dunne, 2007, p. 33; Childress, 2007, p. 44; Lommel, 2005, p. 9). Translators must carry out terminological searches but these take a lot of time. The amount of time invested in terminological searches will affect a translator's productivity and, in the end, this will be reflected in the freelancer's or translation service's bottom line. Properly documenting previous searches accelerates translators' work if the term appears again in a text, saving time and money.

The remaining benefits and risks to be avoided will also be of definite interest to the translator. Obviously, a translator has an interest in producing clear and accurate documents that facilitate client-company communications, strengthen a company's brand and reduce any liability risks related to potential misinterpretations. These are all by-products of a quality translation and all contribute to improving a translator's reputation and securing his or her clientele and business success.

2.1.4 Obstacles to terminology management

In spite of the significant number of reasons to manage terminology, there are still those who do not engage in this practice. The most common reason given for not carrying out terminology work is the lack of time and budget (Lommel, 2005, p. 2). Therefore translators, terminologists and terminology advocates across the world face the challenge of proving to their managers or to themselves that the initial investment will reduce costs over

¹⁰ The exception to the norm is contracts where the translator or translation service charges for revision by the hour.

¹¹ Depending on the country, it is also common to see translation charged per character, line, page or hour.

time: in financial terms, this means proving that terminology work has a good return on investment (ROI).

This challenge is shared by a number of participants in the community: one of the goals of LISA's Terminology SIG was to “[d]etermine and promote the economic value of managing terminology” (LISA, 2008). This topic has been the subject of research by the Translation Bureau in Canada (Champagne, 2004b) as well as by other experts in the field (Dunne, 2007; Childress, 2007).

Establishing the ROI of terminology management is not a simple task, mostly because terminology as a service is usually billed as part of the overall translation process and is not broken out as a separate cost (Champagne, 2004b, p. 9). In some cases carrying out this kind of evaluation has been considered to be too time- and labour-intensive, and in others the benefits of terminology management have been considered to be obvious to its direct users (Kelly and DePalma, 2009, p. 13).

However, experts have found ways of estimating the ROI of terminology management by means of reported studies and focus groups (Champagne, 2004b); calculations of time invested in creating a record, time invested in term searches during translation and the number of times a single term record is looked up (Champagne, 2004b; Warburton, 2008; Kelly and DePalma, 2009); and comparisons of the costs of creating term records with the potential consequences of not doing so (Childress, 2007, p. 44; Kelly and DePalma, 2009). Indirect evidence can also be gleaned from studies in related fields, such as the evidence that implementing tight terminological control on a source text before sending it out for translation can reduce revision costs – particularly for texts that are translated into multiple target languages. For instance, as noted by Brown (2003, p. 4):

... an error requires one hour to fix in the English source, the content is being translated into 34 languages, and it costs \$50/hour for the localization engineer to fix it. That one error has added 34 hours to the localization schedule and will cost \$1,700 to fix. If that same error is caught and fixed before it goes to localization, the error takes one hour to resolve and costs \$50.

These investigations (Brown, 2003, p. 4; Champagne, 2004a, p. 37; Warburton, 2008; Childress, 2007, p. 44) all support the conclusion that it is worth investing in terminology management.

2.2 Terminology management: the how

As previously mentioned, there is little in the way of literature addressing terminology management for translators, let alone terminology management for translators using TEnTs. A translator looking for guidance on a methodological framework to build his or her TEnT terminology database must rely on scarce and brief comments in translation textbooks and on the wealth of theoretical and practical works from the field of terminology that describe how to design and compile a specialized dictionary or terminology database or carry out ad hoc research from a terminologist's point of view.¹² This section will present an overview of the body of recommendations on how to manage terminology that have been gleaned from translation textbooks and other translation literature as well as from the terminology literature.

¹² Owing to the lack of academic or industry literature available on this topic, translators turn to more experienced TEnT users for advice. Therefore, informal sources such as online discussion boards are also a source of information for many translators – particularly since few other options are available to them – but such sources are anecdotal at best and so have not been included in this formal literature review.

2.2.1 Terminology management in the translation literature

Given the number of translation textbooks on the market, the review of this type of literature does not attempt to be exhaustive. Instead, a few examples will be presented to illustrate the guidelines for terminology management that translators can typically find in this type of work. Included in this overview are general translation textbooks, professional translation textbooks and a case study report.

Firstly, general translation textbooks, such as *La traduction raisonnée* (Delisle, 2003) or *La traduction de l'anglais au français* (Ballard, 2002) may not address the issue of terminology management at all. Others, such as Newmark (1988, p. 152) or Robinson (2003, p. 128), offer passing references to the role of terminology in translation without giving any direction as to how translators should translate it or manage it.

In contrast, specialized or technical translation textbooks do tend to address the topic of terminology. However, these types of textbooks will often focus on the role of terminology in the specialized text and the terminological needs of the translator working in such areas. For example, in the textbook *Técnicas documentales aplicadas a la traducción*, M. Teresa Cabré Castellví (1999) discusses the needs of the translator with regard not only to understanding the source-language specialized terminology but also to mastering the equivalent terminology in the target language as well as its syntactic usage, its pragmatic value (degree of harmonization, geographical usage) or phraseological characteristics of the unit (p. 27). Not only does Cabré (1999) describe the terminological challenges a translator may face and the types of reference material that can be consulted (p. 27), she also establishes that translators are often required to carry out multilingual ad hoc terminology research (p. 28). However, in this type of textbook there is no indication as to *how* the translator should carry out this research and, most importantly in the context of this project, how he or she

should record or manage the product of the research.

Secondly, some textbooks that have a more pragmatic approach to the translation industry and are less geared to the linguistic translation process do discuss not only the relevance of terminology management for translators but also the compilation and organization of terminology glossaries or databases. For example, Geoffrey Samuelsson-Brown (2004) in *A Practical Guide for Translators* recommends building a glossary for each translation project in the form of a bilingual list of terms (p. 84). He notes that this list can be compiled manually or with the help of an extraction tool and recorded in a word processor or a TEnT, as long as the lists can be sorted alphabetically and the order of languages reversed (Ibid., p. 85). These glossaries can then be added to a terminology database with each term linked to the client that requested that project (Ibid., p. 86). As to what units should be recorded, Samuelsson-Brown points out that he personally tends to record unknown items (Ibid., p. 109).

For more specific details on how translators should organize terminology, Morry Sofer (2006) offers a more in-depth account of terminology management for translators in *The Translator's Handbook*. With regard to a storage medium, Sofer does not favour a specific tool but stresses that the repository of the terminology information should be easily retrievable and scalable (Ibid., p. 95). As to how to structure the termbase, the classification recommended is to group term records by client as well as by domain and sub-domain, in order to be able not only to identify the subject field in which a term is used but also to easily view what term choices were used in a client's project (Ibid., p. 96). Moreover, the author also recommends taking note of clients' preferences or corporate style, as well as paying special attention to the recording of acronyms and initialisms and whether or not these need to be translated (Ibid., p. 97).

Last but not least, Bert Esselink (2000) in *A Practical Guide to Localisation* offers a holistic and detailed description of how to manage terminology for a localization project. In such a multilingual and electronic resources-oriented field, the recommended medium for managing terminology is an integrated TMS. According to Esselink (2000, p. 398), the basic requirements for such a tool are: flexibility for storing terms, equivalents and supporting information; retrievability of terms with quick and fuzzy search features; and automatic replacement (pretranslation) of terminology in the translation environment. Regarding what units should be recorded, Esselink points out that translators should create their records with terms from three different types of glossaries: operating environment glossaries, client glossaries and project glossaries (Ibid., p. 398).

Firstly, operating environment glossaries¹³ include the terminology of the platform on which the software runs. Secondly, client glossaries contain product terminology standardized across the company or legacy terminological work developed during the life of the company. Finally, project glossaries are to be compiled if no previous version of the software or website has been translated. These project glossaries are built from previously translated product documentation, help files, and technical documentation, upon which a terminology extraction tool may be run to extract the key industry terms (Ibid., p. 400). Other than that, such glossaries will also contain help file titles, manual titles, chapter titles, commonly used verbs or phrases and core software interface items (e.g. menu names, options in dialog boxes) (Ibid., p. 401).

¹³ Operating System providers offer glossaries to help software developers integrate with their tools. For example, Microsoft offers glossaries to developers (<http://support.microsoft.com/kb/140764>) or to the general public (<http://www.microsoft.com/resources/glossary/default.msp>), and so does Apple (<http://developer.apple.com/internationalization/download/>).

Esselink also provides advice on how to structure term records, which he indicates should be concept-based, i.e. with a record for each concept with all synonyms and equivalents for that concept on the same record (Ibid., p. 399). Each record should contain “the terms and phrases associated with the concept and other information such as definitions, target language equivalents, grammatical information on terms, and contextual information” (Ibid., p. 399). In addition, particularly when working on software localization, it would be advisable to include in the record hot keys/shortcuts, the product name and version where the feature appears, as well as the location in the software (button, menu, dialog box title, etc.) (Ibid., p. 403).

All three authors (i.e. Samuelsson-Brown, Sofer and Esselink) stress the importance of terminology management for ensuring the consistency and quality of translations. Note the fact that in all of the textbooks presented above, the classification of terminology by client or translation project is regarded as essential. Moreover, in the case of Esselink, where terminology management is discussed specifically within a TEnT, it is especially interesting to see how the units to be recorded are not limited to key industry concepts or unknown items, but also include “[w]ords or even phrases that are repeated throughout the project”, “[p]roduct-related names that are not to be translated” and “the software user interface strings” (Ibid., p. 403).

2.2.2 Terminology management in the translation technology literature

In *Electronic Tools for Translators*, Frank Austermühl (2001) not only describes the nature of TMSs, but gives a detailed step-by-step overview of how to create a terminology database – in his particular example, with MultiTerm (Ibid., p. 109). He describes the types

of information that can be collected and classifies them in three groups: administrative data (e.g. author, date of creation, date of modification, project, client, etc.), encyclopaedic data (e.g. definition, images, domain, etc.) and linguistic data (e.g. gender, part of speech, context, collocations, etc.) (Ibid., p. 110). Austermühl recommends creating a term record structure whose level of detail will vary according to its purpose (e.g. he proposes that company wide databases require more administrative information than personal ones) (Ibid., p. 110). Finally, he describes the types of fields available within MultiTerm. With regard to what units should be recorded in a termbase, this topic is not directly discussed, although he refers to “terminology” (Ibid., p. 102) and “terms” (Ibid., p. 110) as the units that will be stored in the database.

Lynne Bowker (2002a) in *Computer-Aided Translation Technology: A Practical Introduction* reviews the type of computer tools available to assist translators with their translation tasks. TM systems and TMSs are part of the inventory covered. The purpose and different functions of these tools are described and analyzed to pinpoint their advantages, disadvantages and implications. As part of this discussion, Bowker indicates that it is known that in the localization industry glossaries often include only source and target terms (2002a, p. 87). This is attributed to market demands rather than to the tools being used. However, she also brings up the fact that users have also been observed to have started recording not the canonical form of terms but the most frequent form as well as frequent phrases or expressions that are not necessarily terms in order to benefit from the one-click insertion feature available in TEnTs (Bowker, 2002a, p. 88).

C.K. Quah (2006) in her work *Translation and Technology* presents a chapter on TM tools and TMSs (Ibid., p. 93). She introduces the different tools that integrate such systems, their matching logic and types of matches. With regard to how to structure a termbase or

what content to add, Quah observes that records are classified by concept and not by form – i.e. recording all denominations of a concept in a record and creating one record for each meaning of a polysemous term (Ibid., p.105). According to this author, term records may include additional information such as definitions, contexts, gender and synonyms (Ibid., p. 106).

Marie-Claude L’Homme’s *Initiation à la traductique* (2008¹⁴) describes in detail the nature and the functioning of database management systems (Ibid., p. 119) – including TMSs (Ibid., p. 32) – and automated lookup tools – covering TM systems (Ibid., p. 174) and active terminology recognition (Ibid., p. 182). She points out that, although the most popular unit to be recorded by translators will undoubtedly be the term, translators may also be interested in recording polysemous words, fixed or frequent expressions as well as parts of or whole sentences (Ibid., p. 146). Finally, L’Homme indicates that these units will be accompanied by supporting information which includes but is not limited to definitions, equivalents, notes and contexts (Ibid., p. 146).

2.2.3 Terminology management in terminology literature

Given that neither the translation nor the translation technology literature offers an in-depth discussion on how to carry out terminological tasks or how to build terminology databases for translators, these language professionals may turn to the wealth of literature on this topic found in the field of terminology. However, as noted above, it is important to keep in mind that the terminology literature is almost exclusively targeted at terminologists, who have different priorities, goals and needs than translators.

¹⁴ The same content can be found in the first edition of the book published in 1999.

Terminology, as seen in chapter 1, is defined by Juan C. Sager as “the study of and the field of activity concerned with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialised areas of usage of one or more languages” (1990, p. 2).

The purposes of terminology as a discipline have evolved over time. While Wüster viewed the purpose of terminology almost exclusively as the standardization of concepts and denominations (Cabré, 1999, p. 111), more recent perspectives on terminology claim a more descriptive approach. Such is the case of Cabré’s CTT (Communicative Terminology Theory, in Spanish, *teoría comunicativa de la terminología – TCT*), whose goal is to “collect terminological units within a specific subject and situation and establish their characteristics according to this given situation” (1999, p. 124, my translation¹⁵). L’Homme shares a similar vision for terminology: its purpose is to describe the terms used within a specialized subject field (2004, p. 21) and the criteria to identify terminological units are based on the units’ lexical and semantic relationships (Ibid., p.32).

2.2.3.1 Terminology theories and approaches

Although relatively young, the discipline of terminology has produced a theoretical model that provides a widely adopted basis for collecting terminology and producing terminological tools such as glossaries, specialized dictionaries or term banks. As seen in section 2.1.1 above, terminologists may practice thematic or ad hoc terminological research. The basic principles of terminology apply to both approaches. A detailed description of the particularities of ad hoc terminological research will be presented below.

¹⁵ Original quote in Spanish: “El objetivo de la terminología aplicada es el de recopilar las unidades de valor terminológico en un tema y situación determinados y establecer sus características de acuerdo con esta situación.” (Cabré 1999, p. 124)

Wüster laid the foundation of terminology theory, which he developed during the writing of the dictionary *The Machine Tool*, the subject of his doctoral thesis in 1938. Later on, he further elaborated this basis into what is now generally known as the *General Theory of Terminology (GTT)*. This theory revolves around principles that will guide the applied terminological method. Cabré (1999, p. 111) summarizes the main ones as follows:

- a) The objects of study of terminology are specialized terms that belong to one field of specialization.
- b) Terms are semiotic units consisting of a concept and a denomination.
- c) Terminological research is onomasiological. A terminologist first identifies a concept in a field of specialization and then looks for its denomination in one or multiple languages.
- d) Concepts belonging to a single sub-field establish various relationships that form a conceptual structure or system.
- e) Terminology studies terms and their relationships in order to standardize concepts and denominations.
- f) Terminology aims at ensuring precision and univocity¹⁶ in professional communications.

¹⁶ Univocity occurs when a term denotes only one concept and that one concept can only be represented by that one term. Terminology, when prescriptive, strives to establish and encourage univocity. However, synonymy (one concept, two terms to represent it), polysemy (one term denoting two concepts) and homonymy (two terms that happen to share the same written form and denote two separate concepts) are often present, even in specialized texts.

These principles led to the development of an applied method consisting of different phases:

- a) Delimiting and defining the scope of the terminological work, including subject field, users and purpose;
- b) Becoming familiar with the subject field and collecting reference documentation;
- c) Extracting candidate terms from the corpus of reference documentation;
- d) Building a conceptual structure of the field¹⁷;
- e) Pruning the list of candidates of terms that do not belong to the field and filling in gaps brought to light by the conceptual structure; and
- f) Creating terminological records for each term that must follow a strict structure for which sample templates have been established to collect the required information for each term: term unit, context, definition, observations, etc., all of which must be linked to the sources from which they were extracted.

These steps will then be repeated for all the languages that the glossary or terminology database is intended to cover. Cabré (1992, 1998), Dubuc (2002), Pavel and Nolet (2001) and the Translation Bureau (2008a) offer detailed descriptions of how to execute each phase of the terminological work.

The main difference between thematic and ad hoc research resides in the number of terms that constitute the object of the research. Ad hoc terminological research aims to solve specific difficulties with one concept or a very small number of concepts (Cabré, 1992, p. 319; Dubuc, 2002, p. 41; L'Homme, 2004, p. 46). For ad hoc research, users (translators, specialists, etc.) send queries to terminologists to identify the term that denotes a concept, the meaning of a term, the equivalent of a term in a specific language, etc.

¹⁷ Phases c) and d) do not have clear delimitations and often overlap.

The difference in purpose of this type of terminological research (i.e. not attempting to establish the terminology for the entirety of a specialized field) affects the method used for the work, which is nevertheless entirely based on the principles of terminology research described above. Cabré (1992, pp. 325-329) and Dubuc (2002, pp. 42-44) describe how to carry out this type of research. The main points of deviation from thematic research are that

- a) research stems from a user query about a specific concept or term or set of concepts or terms;
- b) terminologists will most likely not be able to build the corpora of reference documentation for each query and will instead use general and specialized dictionaries, reference works such as encyclopaedias and specialized publications as well as searches on the Web to this end; and
- c) the resulting product is an individual record or small number of related records but never a full glossary.

Of the two types of terminological research, ad hoc terminology research comes much closer to the terminological needs of translators, as translators will often be the users of such services (Cabré, 1992, p. 320; Dubuc, 2002, p. 41). On a daily basis, any translator faces the challenge of not knowing a term's meaning or equivalent. In such circumstances, translators carry out the first steps of ad hoc research and, if results are not found easily in existing resources, may call upon the services of a terminologist (if available) or carry out the research themselves in full (Jaekel, 2000, p. 163) or in part (Joscelyne, 2000, p. 91).

2.2.3.1.1. Challenges of the GTT

In spite of the widespread praise of the GTT's virtues – namely its systematicity, logic and efficiency for achieving standardization of denominations in specialized fields – GTT faces criticism on the theoretical front and requires adaptation to the world of possibilities now offered by new technologies.

GTT critics firstly disapprove of its purpose being limited to the standardization of denominations in specialized domains, as over time the field of terminology has evolved and terminologists have recognized that not all terminological work must be prescriptive. Secondly, GTT's principle of term univocity has proven to be more of an ideal as the presence of variation, even in highly specialized fields, is irrefutable and in some cases deliberate and useful. Thirdly, GTT perceives terms as concept-denomination pairs and sees these as being different than lexical units. Critics tend to perceive terms as words and therefore as natural language. Although they may differ as to how or when a word acts as a term, they agree that terms are words and that how they are integrated into text is a very important aspect that does need to be recorded. Finally, another of the most criticized aspects of the GTT is that it asserts that concepts are denoted primarily or even exclusively by nominal forms. Those who see terms as part of natural language tend to disagree with the idea that terms belong exclusively to the noun category and recognize verbs, adverbs, adjectives and phraseological expressions as categories able to function as terms and suitable for recording on term records.

Those who have criticized and proposed alternatives to the GTT have done so from different perspectives: linguistic, cognitive and communicative (Sager, 1990; Cabré, 1999), sociocognitive (Temmerman, 1997, 2000), text linguistic (Bourigault and Slodzian, 1999), sociolinguistic (Gaudin, 2003) and lexico-semantic (L'Homme, 2004).

New technologies have also challenged the method proposed by Wüster. Advances in the computer science world since the mid-20th century have been innumerable: the personal computer, word processors, databases and syntactic parsers, text analyzers, bilingual term extractors, knowledge bases, search engines, etc. These innovations have put an array of tools in the hands of terminologists to enable them to better work with text. Terminologists have so eagerly adopted this new way of working that they even coined a term (*terminotics*)¹⁸ to describe those terminology tasks that involve the use of computer software (L'Homme, 2004, p. 17).

Terminologists quickly realized that computer programs can help to automate practically every single phase of terminology work and to process texts at speeds unimaginable for the human brain, allowing corpora of reference documentation to grow to previously unthinkable sizes. Accounts of possible applications of computers in terminology date back practically to the arrival of the first personal computer and have continued steadily since that time (Auger, 1989; Cabré, 1992; L'Homme, 2004).

The arrival of all these applications did not necessarily contradict the GTT, but a new theoretical reflection was required to describe the tools in order to better understand them and provide a methodology on how to integrate them into terminology work. As a response to these new needs, we see works such as Bowker and Pearson's (2002) *Working with Specialized Language: A Practical Guide to Using Corpora* and L'Homme's (2004) *La terminologie : principes et techniques* being added to the corpus of terminology literature.

New technologies have, however, underscored what many critics had already pointed out: in spite of the principles of the GTT, terminology work is more semasiological than onomasiological because it almost invariably stems from texts (Bourigault and Slodzian,

¹⁸ This term is even more prevalent in its French equivalent (*terminotique*) and among francophone authors.

1999; L'Homme, 2004, p. 30). New technologies have allowed terminologists to exploit and analyze large amounts of textual data. This new ease of accessing large amounts of text and of manipulating such data has underlined the central role the corpus of reference documentation plays in terminology. It is for this reason that works such as “Pour une terminologie textuelle” by Bourigault and Slodzian (1999) have entered the terminology scene claiming that terminology should be anchored within text linguistics.

2.2.3.2 Applicability of terminology theory to translators

Translators do not carry out terminological work with the goal of describing or standardizing the terminology of an entire subject field. Translators perform terminological work from a problem-solving perspective. A study by Estopà (2001) clearly illustrates this difference. The study presented subject experts, terminologists and translators with the task of extracting terms from a series of texts. Faced with the same challenge, translators identified units presenting a difficulty in their translation or unknown units, subject experts highlighted the items denoting key concepts, while terminologists executed an exhaustive extraction of all specialized units. Based in part on these observations, we can say that a translator's main purpose when designing and building a termbase as a translation aid would be to record all units – whether or not they meet the strict criteria for “termhood” – that pose an obstacle to translation and require a certain amount of research that the translator wants to avoid having to repeat in the future.

Translators do not look for the same kind of information as other users when looking up terms. For example, Durán Muñoz (2010, p. 10) established through a survey that professional translators require linguistic and pragmatic information but do not consider semantic and grammatical information to be essential.

In the case of integrated termbases built with the intention of assisting in the translation process, differences between a terminologist's approach and a translator's approach go beyond term extraction criteria. Given that termbases are an essential component in a TEnT from which the system can automatically retrieve term matches to be inserted in the target text, it will be in the translator's best interest to build the termbase in such a way as to optimize automated retrieval of information based on the form of the term as it appears in the source text. This will influence the nature of the units recorded, their recorded forms and where on the record information will be located.

Several scholarly articles describe the influence of the TEnT environment on termbase design based on anecdotal evidence. For example, Kenny (1999) and Bowker (2011) describe how translators seem inclined to record expressions such as slogans, formulas, or even phone numbers, which would not fall under the traditional concept of a terminological unit. They may also choose to record information that would not usually appear in a term record, i.e. all inflected forms of a term, collocations, hyperonyms and so on. O'Brien (1998, p. 118) notes that translators tend to create bilingual lists of terms rather than term records with multiple fields. Bowker (2011, p. 221) points out that translators seem willing to disregard the traditional "ban" on using translated documents as reference material for term equivalents and that they may be more prone to organize their termbases semasiologically (by form) rather than onomasiologically (by concept) (Ibid., p. 223).

In short, although translators can use terminology theory as a basis for conducting terminological research, they will have to adapt the terminological method to serve their own purposes: to document their searches and optimize the performance of their TEnTs. The current research attempts to explore how translators are carrying out this task and to evaluate different strategies in order to determine if there is an optimal one.

2.2.3.3 Sample bilingual record structures found in terminology literature

Translators can also resort to terminographical works for guidance on how to practice terminology. In such works they will find proposed record structures that they can use as reference for their integrated-termbase record templates. Here are some record structures proposed by several authors. Fields in bold are considered mandatory.

- 1) Bilingual record structure proposed by Dubuc (2002, p. 85)

L1 term; synonyms; abbreviations

Source
Source date
References
Grammatical observations
Usage and logic observations
Context

L2 language term; synonyms; abbreviations

Source
Source date
References
Grammatical observations
Usage and logic observations
Context

Domain

Author

- 2) Bilingual record structure proposed by Pavel and Nolet (2001, p. 48)

L1 language preferred term(s)

Synonyms

Abbreviations

Spelling or syntactic variants

Quasi-synonyms

Definition
Context
Observations

L2 language preferred term(s)

Synonyms

Abbreviations

Spelling or syntactic variants

Quasi-synonyms

Definition
Context
Observations

- 3) Bilingual record structure proposed at the *Le grand dictionnaire terminologique* (OQLF, 2002)

Domain

Sub-domain
Domain of the sub-domain
Main Term
 Part of speech
Officialization
Definition and Note
Sub-entries
 Synonym
 Quasi-synonym
 Feminine Form
 Abbreviation
 Graphic variation
 Transcribed form
 Loan
 Non-retained term
 Term to avoid
Illustration
Equivalent
Author

2.2.4 Terminology management in practice

During the last decade, a series of studies have been carried out, in different translator groups, on the use of TM software and on terminology management in particular. I have relied on these to introduce the current status of terminology management within the translation industry; they will be essential to contextualize the results of my own survey.

Appendix B presents a brief description of each survey that indicates its objectives, target audience, respondents' profile, means of distribution, etc. The details of the results obtained in each survey will be discussed in parallel with the results of this current research project. The surveys are presented in chronological order in Table 2 below.

Year	Title	Author
2002	<i>LISA Translation Memory Survey: Translation Memory and Translation Memory Standards</i>	Arle Lommel
2003	<i>Translation Memory Survey</i>	Mary Höcker
2003	<i>A Major Breakthrough for Translator Training (eCoLoRe)</i>	Alan Wheatley
2004	<i>Portrait of Terminology in Canada</i>	Guy Champagne
2004	<i>The Economic Value of Terminology: An Exploratory Study</i>	Guy Champagne
2004	<i>LISA Translation Memory Survey: Translation Memory and Translation Memory Standards</i>	Arle Lommel
2005	<i>LISA Terminology Management Survey: Terminology Management Practices and Trends</i>	Arle Lommel
2005	<i>ATIO Survey of Independent Translators</i>	Nancy McInnis and Maha Takla
2005	<i>Translation and Technology: A Study of UK Freelance Translators</i>	Heather Fulford and Joaquin Granell-Zafra
2006	<i>OTFLAQ Survey on Rates and Salaries</i>	François Gauthier
2006	<i>Translation Memory Survey. Translation Memory Systems: Enlightening the Users' Perspective</i>	Elina Lagoudaki
2006	<i>eCoLoTrain Results. Translator Training Survey</i>	eCoLoTrain
2006	<i>Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption</i>	Sarah Dillon and Janet Fraser
2007	<i>Translation Memory Survey</i>	Institute of Translation & Interpreting (ITI in UK)
2007	<i>ATIO Survey of Salaried Translators</i>	Association of Translators and Interpreters of Ontario (ATIO)
2008	<i>OTFLAQ Survey on Rates and Salaries</i>	François Gauthier
2008	<i>On the Lighter Side: Terminology Results (ATIO)</i>	Nancy McInnis
2009	<i>The Case for Terminology Management</i>	Nataly Kelly and Donald A. DePalma
2009	<i>Terminology: An End-to-End Perspective</i>	SDL Trados
2010	<i>Specialised lexicographical resources: a survey of translators' needs</i>	Isabel Durán Muñoz

Table 2 List of Previous Surveys on Terminology Management and TEnT use

3 Preliminary hypotheses

Two observations act as springboard for the hypotheses of this project:

Observation 1: There are no generally accepted best practices on how to design and manage a translation-oriented terminology database integrated with a TEnT.

Observation 2: The closest body of literature that could serve to guide a translator on how to design and manage a translation-oriented terminology database integrated with a TEnT is the terminology and terminography literature.

These two observations lay the groundwork for an overarching conceptual hypothesis:

Translators working interactively or pretranslating with TEnT termbases adapt the terminographical method to their specific needs: active term recognition, one-click insertion in the translated text and pretranslation.

As presented in section 2.2.1, literature on how translators can best design and populate an integrated termbase is scarce. Translators can turn to terminology and terminography literature for basic information on how to manage their terminology within TEnTs (see section 2.2.3). However, the needs and goals of a translator working within a TEnT differ from those of a terminologist carrying out systematic or ad hoc terminography (see section 2.2.3.2).

Translators do carry out ad hoc terminological research and terminography to better understand or find an equivalent for problematic terms or to coin a translation in the target language for state-of-the-art concepts that have not yet been recorded in reference sources. In such cases, they must turn to terminology theory and terminography principles to guide them.

Nevertheless, if these were the only types of units recorded in their integrated termbases, translators would be greatly under-using this type of tool; they would be treating it as if it were a stand-alone terminology database. Such an approach would disregard the added functionalities of integrated termbases: active term recognition, one-click insertion of equivalents, and pretranslation and term record creation capabilities at any point in the translation workflow. (For a more detailed explanation of these functions, see section 1.1.1.) They would also be ignoring the essential purposes of the TEnT and its integrated termbase: to facilitate the translation of a text, to automate as much as possible term recognition and equivalent insertion during the translation task in order to maximize the use of the records created, and to encourage consistent use of terminology.

Therefore, terminology theory and terminography principles alone will not meet all the needs of a translator designing and using an integrated termbase. Taking this difference of purpose into account, based on the literature review and personal experience, we present a set of sub-hypotheses that will contribute to the design of a termbase geared to facilitate and improve interactive translation within a TEnT.

- a) Contrary to what current terminology and terminography literature recommends, translators will use fewer term record fields in a TEnT-integrated termbase.
- b) Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard.
- c) Contrary to what current terminology and terminography literature recommends, translators will classify records in personal TEnT-integrated termbases first by client or project and only secondly by domain.
- d) Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.
- e) Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.
- f) Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will not be opposed to extracting terms/units and equivalents from translated texts.
- g) Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or their most frequent form(s).

3.1 Sub-hypothesis A:

Contrary to what current terminology and terminography literature recommends, translators will use fewer term record fields in a TEnT-integrated termbase.

In terminology theory, a term record may document a great amount of information on a unit, given that the goal of creating that term record is to inform potential users of the nature of the concept recorded, how this concept is integrated into its domain of specialization, what the terms that denote this concept are and, if applicable, the connotations, degree of correctness/acceptability and geographical usage of each equivalent. Term banks are available to a variety of potential users (language specialists, subject specialists, subject-field students and the general public) with very different needs, and the diverse content of the banks is a reflection of the range of users and uses.

According to Sager (1990), the information recorded can be administrative, bibliographical and terminological. This last category may in turn be subdivided into conceptual, linguistic and pragmatic information. The actual number of categories that can be recorded is immense. ISO 12620:2009 *Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources* establishes an internationally accepted set of about 200 data categories that can be used in a terminology database, among other linguistic resources.

The above ISO standard is an exhaustive list of possible data categories aimed at providing a category for any type of information that may need to be recorded. It by no means intends that all categories should be used in a single termbase, much less in all of them.

ISO 30042:2008 *Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)* proposes a terminological mark-up language in order to standardize information

recording and exchange. TBX establishes a number of default categories which, although they do not include all categories listed in ISO 12620, still include several dozen types. A terminology database structure does not need to use all default categories to be TBX-compliant, but it must use default categories as established in the standard¹⁹.

Terminology manuals recommend a basic term record structure with a number of specific fields. For example, Dubuc (2002, p. 83) recommends that a term record include the following fields: entry term, source, source date, references, grammatical information, usage information, context, domain, author and date²⁰. Another example is Pavel and Nolet's (2001, p. 9) recommendation that at the very least a term record must "inform the user about the subject fields of the concept, the languages in which the concept is described, the terms that designate the concept in each of these languages, the definition of the concept (or any other type of textual support), and the sources that document this information." Textual supports according to Pavel and Nolet (2001, p. 49) include definitions, contexts, observations and phraseology. We have seen illustrations of the structures recommended by Dubuc (2002, p. 83) and Pavel and Nolet (2001, p. 48) in section 2.2.3.3.

While these record structures are ideal for terminological resources created by terminologists targeting language and subject specialists as their users, translators will rarely create such complete records for themselves (or for a colleague). The reasons for this are numerous:

¹⁹ Additional categories can also be used as long as the nature of the data recorded and its logic in the structure is properly defined within the markup language.

²⁰ When working with paper term records, Dubuc also includes a field for keywords where the author would indicate synonyms, related terms and term descriptors and for each of which an additional record cross-referencing the main record would be created (2002, p. 83). Dubuc points out that keywords are not necessary for electronic records because all words of an electronic record become keywords (2002, p. 83). This also only applies if all fields of the electronic record are searchable.

- Translators carry out terminological work in order to complete a translation project²¹, which usually means that such work is done with a very short deadline.
- Translators may work on very varied topics or cutting-edge fields where terminology evolves and goes out of date fairly quickly.
- Translators may omit information if their terminology databases are created for their own use or to be shared with a small group of users if they deem that the information is obvious or generally known.
- Translators are not generally formally requested or paid to carry out terminological work, which may discourage them from investing a large part of their time in creating term records.
- Translators can now carry out term research much faster thanks to online resources and tools (e.g. Internet search engines).

The fact that translators and localizers require a limited number of fields in their records was acknowledged by LISA, which developed the TBX-Basic mark-up language specifically geared to meet the localization industry's needs (LISA, 2009, p. 4). LISA studied the terminological needs of the localization industry by means of a series of surveys and established a basic term record structure that includes as mandatory fields only the term entry, its language code and the field "part-of-speech" if the terminology database is meant to be machine processed. If it will be used exclusively by humans, it must include either a definition or a context. Using at least one category field is highly recommended, with the domain or subject field being the most popular. TBX-Basic includes a series of other fields that may be used optionally. For a more detailed description of the record structure

²¹ Translators working in-house may have the additional mandate to create or contribute to a company-wide glossary or terminology database. Under such circumstances their approaches may differ as the terminology database may not be used exclusively for translation.

proposed in TBX-Basic, please refer section 5 of the Integrated Termbase Optimization Survey questionnaire, which can be found in Appendix D.

These reduced needs of the localization industry are also confirmed by authors in the field of translation. Samuelsson-Brown (2004, p. 84) recommends building glossaries consisting of bilingual lists of terms, and O'Brien (1998, p. 118) notes that translators often create glossaries that contain only terms and their equivalents.

3.2 Sub-hypothesis B:

Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard. We have already pointed out that there is a lack of guidelines on how best to organize a TEnT-integrated termbase and that translators require less information for their own termbases than terminologists do for term banks. However, if translators were provided with an industry-approved template that they could modify to meet their own needs and that came readily available within their TEnT, we believe that they would likely opt for creating their termbases based on this model. Translators would not fill out all fields systematically, but when they found a piece of information worth recording, they would enter it in the appropriate field.

We are confident that such an approach would be well accepted by the translation community for three main reasons:

1. The lack of guidelines on how to design and build a TEnT-integrated termbase makes the initial set-up process complex and stressful for translators. Starting off with a poorly designed template structure can mean a life sentence of working with an

awkward database or long hours of modifications, data exporting and re-importing at a later date. Having a TEnT tool with a built-in template that they can select and minimally modify (changing field names, adding or removing fields, etc.) – along with a guarantee that the tool will warn them if their modifications go against the TBX-Basic standard requirements – would give translators a good starting point for building their termbase. This would save translators time in deciding which fields can be usefully included, what level of the record each field should be placed at, which fields should be mandatory, etc. Translators could still personalize their termbases, but they would not be starting off with a blank slate.

2. The fact that this readily available template would be based on an industry standard and not simply on the software provider's best practices would instil confidence in the reliability of the template and encourage the translator to use it.
3. The standard is being promoted and it is on its way to becoming the termbase standard of reference. Translators who are considering the possibility of sharing their termbases at some point, or who have experienced migrations from one TEnT to another, will be inclined to use a template that ensures that their termbase follows a standard in order to facilitate the exchange and transfer of their termbases in the future.

Given that this sub-hypothesis is a prediction of how users will react to a hypothetical scenario, it could not be included in the first survey of this project, which addressed actual current terminology management practices within TEnTs. This sub-hypothesis will be addressed in the second survey, which focused on user-acceptance of certain guidelines.

3.3 Sub-hypothesis C:

Contrary to what current terminology and terminography literature recommends, translators will classify records in personal TEnT-integrated termbases first by client or project and only secondly by domain.

In terminology theory, classifying records according to a domain and even sub-domain is key, as a main goal of terminology work is to establish or describe the terminology of at least one and often several specialized domains. Therefore, a domain field is present in all term record structures recommended by terminologists (e.g. Dubuc, 2002, p. 83; Pavel and Nolet, 2001, p. 9).

As Estopà (2001) points out, translators are interested in unknown terms or units that present a challenge for translation. Another main concern in translation is to ensure that terminology is used consistently for each client and that a client's preferred terminology is applied correctly. For these reasons Esselink (2000, p. 398), Samuelsson-Brown (2004, p. 84, 86) and Sofer (2006, p. 96) indicate that, in the context of translation, records must first and foremost be classified by client or project. However, this does not preclude the option of also classifying the terms by domain and/or sub-domain (Sofer, 2006, p. 96).

While client and project classification are easily identifiable facts that can be semi-automatically filled out in certain tools, classifying a term according to its specific domain and sub-domains can be laborious and time-consuming. Often such a decision may not be clear-cut and may require extensive reading on the subject matter. This may discourage translators from using such a classification. Alternatively, translators may use broader domain categories than terminologists to facilitate the classification task.

The drawback of not classifying terms by domain is that it makes a termbase much less shareable. Unless the termbase is shared with a translator working in the same field of expertise or with the same set of clients, the client and project classification fields will

provide very little additional information, while a domain field will be much more useful for helping any new user to navigate and use the termbase. For this reason, this sub-hypothesis applies to personal termbases and not to shared ones.

3.4 Sub-hypothesis D:

Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.

One of the key principles of traditional terminology is that records are created following an onomasiological or concept-based approach (i.e. one where each record represents a single concept and all denominations of a concept are entered in the same record) (Cabré, 1998, 30; Dubuc, 2002, p. 28; L'Homme, 2004, p. 26; Translation Bureau, 2008a). Indeed, the fact that terminology starts with the concept, from this arrives at its denominations (i.e. the terms), is traditionally one of the key differentiators between terminology and lexicology. The latter works semasiologically starting from a denomination or form and aiming to establish all its meanings or all the concepts it denotes (Cabré, 1998, p. 30; Dubuc, 2002, p. 27; L'Homme, 2004, p. 23).

This methodological approach underlies the way that terminology resources are organized. In a lexicographical work such as a dictionary, there is generally one entry per lemma under which all meanings will be listed (in a form-based or semasiological approach). In contrast, in a terminographical work each concept will have its own entry, which will list all forms denoting that one concept (in a concept-based or onomasiological approach).

An onomasiological organization is very efficient when one looks up a concept in order to master its meaning and denominations. All relevant information on a concept can

be found in a single record: definition, denominations (i.e. terms, synonyms and variant forms), phraseology, grammatical observations, usage observations, etc.

However, in the context of interactive translation, a key factor in play is the ease of insertion of equivalent terms into the target text. For this purpose, if a unit has synonyms but they cannot be used interchangeably, or if multiple forms of the target term are recorded, it will be more efficient to create separate records for each of the synonyms or forms.

For example, in the province of Québec, municipalities levy a tax on the sale of immovable properties (e.g. houses). Its official denomination is “duty on transfers of immovables” in English and “*droits de mutation immobilière*” in French, but it is more commonly known by its endearing appellation “welcome tax” in English and “*taxe de bienvenue*” in French. In a terminological record, both forms, the official term and the familiar one, would be part of the same record. However, when translating a text containing the term *droits de mutation immobilière*, we would rarely want to replace it with its more colloquial synonym “welcome tax”, and a translator would be distracted if he had to choose between the two forms. Therefore, it would be more practical to create two separate records so that when *droits de mutation immobilière* appears in a text, the TEnT proposes only “duty on transfers of immovables” as a solution, and when *taxe de bienvenue* appears, it proposes only “welcome tax”.

The same will apply when several forms of a term or unit are recorded. When translating a text into French where terms appear in the plural, for example, it will be more efficient not to be presented with feminine and masculine singular forms of the term. Translators who work between two Romance languages may prefer to create separate records for the masculine singular, feminine singular, masculine plural, and feminine plural

forms and so on²². It can become even more crucial to separate the records when it comes to verb forms with the different persons and tenses.

This hypothesis does not exclude the possibility of creating concept-based records if they are univocal or if synonyms are fully interchangeable. It must also be noted that although this approach gives special weight to the form, it must not be confused with the semasiological or form-based approach used in lexicography. In lexicographical works, all meanings of a unit will be listed under that unit regardless of whether its equivalents take different forms. In a terminology database, to encourage ease of insertion, filtering and retrievability, a record describes only a single concept but includes all terms for it. The translators' approach described above thus falls somewhere between the conventional onomasiological and the semasiological approaches.

3.5 Sub-hypothesis E:

Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.

One of the key questions in terminology is the definition of “termhood” acceptance criteria in order to establish whether or not extracted term candidates qualify as actual terms. To reduce the discussion to the bare essentials, a term is a unit designating a specialized concept (Cabr e, 1998, p. 149; Dubuc, 2002, p. 33). Although in principle, this seems like a straightforward distinction, applying it is rather complex. Authors propose various guidelines for separating terms from non-terms. Here are two examples. Firstly, Dubuc (2002) recommends observing the following characteristics:

²² This technique will give its best results when terms share the same gender across languages. When there is a gender change the translator will have to verify and adapt the agreement with accompanying determiners and adjectives.

- a) lexicalization (Ibid., p. 59): the *Handbook of Terminology* defines lexicalization as “The process by which a group of words comes to be fixed by usage and to behave as a single lexical item.” (Pavel and Nolet, 2001, p. 111). While the structure N+N is often an indicator of a high degree of lexicalization (e.g. *appeal process*), usage may also lexicalize other structures (e.g. *Speaker of the House of Commons*)²³;
- b) class or opposition mark (Dubuc, p. 60): in a combination of determiner + determined, the more a determiner serves to distinguish a type of a certain notion or serves to differentiate a notion and its coordinate concepts (e.g. *French bread, Italian bread, Ciabatta bread*), the more likely it is to be a term;
- c) cooccurrence (Ibid.): if a combination of words appears frequently together in a corpus of texts of a specific domain, this expression is likely to be a term (e.g. *file sharing*);
- d) typographical indicators (Ibid.): authors may highlight (bold, italics, underlining, small caps, etc.) key terms in a text.

L’Homme (2004) suggests a different series of guidelines more focused on the semantic aspects of the term candidate:

- a) specialized meaning (Ibid., p. 64): the unit must denote a concept of a specialized domain to be identified as a term;
- b) nature of semantic actors (Ibid.): firstly, the actants²⁴ that the unit refers to and interacts with must also belong to a specialized field; secondly, if the unit keeps the

²³ This is a good example of how the term / non-term border can be rather fuzzy, as some may consider the entire unit a term while others will see it as a collocation of two distinct terms, *Speaker of the House* and *House of Commons*.

²⁴ Actants can be syntactic or semantic. Syntactic actants are easier to recognize as they coincide with the nominals that represent the subject, objects and complements of a verb (Mel’čuk, 2004, p. 2). Semantic actants have the same binding relationship as the subject, objects or complements do to a verb but in the context of a term collocation (Mel’čuk, 2004, p. 7).

- same meaning when combined with non-specialized actants, it is not a term. For example, to *file* in the context of submitting one's tax report to the revenue agency will be an accounting term unit. However when it appears in the context of putting documents away in classified compartments, it will be a non-term unit;
- c) morphological similarity (Ibid., p. 65): once a term has been identified according to the two previous criteria, its derived and inflected forms will by default also be terms (e.g. *filing*);
 - d) paradigmatic relation (Ibid., p. 66): once a term has been identified according to the guidelines presented in a)–c), any units that include the term or relate to it by representing types of this unit, or opposites, will also be terms (e.g. if *propeller* is a term, then *hub* and *blades* that denote parts of a propeller are also terms).

These are only two examples of the strategies proposed to identify terms. None of the guidelines described above are absolute. Other authors present different criteria, and not all authors agree. The two samples above have been provided to illustrate the fact that separating terms from non-terms is an essential requirement for terminology and that it is by no means an easy task.

In a translation-oriented integrated termbase, translators will be interested in side-stepping this limitation to record freely any type of information that either presents a translation difficulty or merely occurs frequently in a text. In both cases, it will be more efficient for translators to record the unit once in order to be able to insert its equivalent quickly in the target texts. Therefore, in a translation-oriented termbase, recorded units will meet at least one of the following criteria:

- They will be units unknown to the translator or be units that are difficult to translate. Estopà (2001) established that these are the units that translators identify as requiring research.

or

- They will be units that appear frequently in the text to be translated and which can be more quickly inserted into the text if recorded in a glossary. One of the functions of an integrated termbase is to look up terms in the database automatically and to allow the user to insert them easily in the text. It is for this reason that translators will be interested in recording not only terms but also slogans, formulaic expressions, telephone numbers, web addresses or simply chunks of text that recur often. Kenny (1999) and Bowker (2011) note in their research that translators are already implementing such practices.

or

- They will be units that present a formal challenge (.e.g. units prone to eliciting typographical errors, units with unusual capitalization, long units). The one-click insertion feature that TEnTs provide can also be used as a type of powerful auto-correct function to quickly insert bits of texts that otherwise would be more laborious to type.

3.6 Sub-hypothesis F:

Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will not be opposed to extracting terms/units and equivalents from translated texts.

In terminology, one of the key principles is that terms and their equivalents must be extracted from authentic texts originally written by experts in the language in which the term is to be extracted. For this reason, using translated documents as references for extracting terminology is not recommended (Dubuc 2002 p. 164; L’Homme, 2004, p. 126). Dubuc (2002, p. 51) accepts as the only exception to this rule cases where specialized documentation written originally in a language does not exist. This principle ensures that the extracted term will indeed be a true representation of the denomination actually used by experts in that specific language.

L’Homme (2004, p. 126) and Bowker (2011) both point out that the principle of not using translated documents as references for terminology work is being increasingly breached. In those cases, L’Homme sets the condition that selected translations must represent real usage in the target language (Ibid.). More precisely, she indicates that terminologists are turning more and more often to bilingual and multilingual corpora of aligned texts (i.e. TMs) (Ibid., p. 140) as one of the advantages of such resources is that they allow translators to identify equivalents more quickly (Ibid., p. 131).

In Pavel and Nolet’s (2001) *Handbook of Terminology*, we see another example of the change in practitioners’ attitude toward the principle of using original-language documents exclusively. The authors recommend “that original-language sources in the source and target languages be scanned for terms first, followed by translated sources” (Ibid., p. 41). We see how in this case translations are not excluded, but they are considered to be second-class

sources. The authors go one step further when it comes to terminology extraction using a department's or company's bilingual documentation. In such cases, they recommend using source documents and their target translations as primary resources (Ibid.). If possible, the term candidates extracted from translated documents should then be verified in original-language reference documents (Ibid.).

Translators, however, may not have the time or the means to access parallel documentation originally written in the target language, to extract term candidates for that concept and to validate them. Therefore, finding an equivalent for a specialized term that is not present in terminological resources such as specialized dictionaries and term banks can be a difficult challenge.

TEnTs offer access to an invaluable resource: the TM database. In such repositories of text, translators can access specialized documents in the source and target language, which are generally aligned at the sentence level, and which they can quickly and easily query to obtain term equivalents that they or other translators have used in the past.

Using such resources as a source for term equivalents has the drawback of not relying on texts written in the target language by subject-field experts: i.e. there is the risk that the translator did not use the most appropriate term for a concept. However, a series of scenarios may lead a translator to turn to translated texts as a reference for a term equivalent:

- a) If the translator translated the text him- or herself and is convinced that proper research was done at the time to establish the equivalent, but the equivalent was not properly recorded.
- b) If the translator has guarantees that all texts entered in the TM have been revised and meet quality standards.

- c) If the translator does not have guarantees about the quality of the translated texts but validates the term equivalent candidates extracted from them.
- d) If the translator is mandated by the client to use the terminology found in their previous translations.

3.7 Sub-hypothesis G:

Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or their most frequent form(s).

In terminology resources, units are recorded in their base forms just as we find them in lexicographical works such as dictionaries. Base forms will be singular (unless they exist only in the plural) for nouns, masculine and singular for adjectives²⁵, and infinitive for verbs (Dubuc, 2002, p. 82).

This is a generally accepted and efficient method of recording terminological and lexicographical information. However, in the context of TEnTs, it must be noted that, generally, retrieval techniques are governed by the principle of character-string matching. In other words, these tools will be able to identify a term record match only if the exact set of characters appearing in the text is present in the terminology database as a term entry. Therefore, by recording only the base forms of terms or units, translators would risk not being automatically presented with the information they added to the terminology database any time that a plural, a feminine or a conjugated verb form appears in the text.

Some TEnTs offer fuzzy term recognition, which allows the TEnT to retrieve a term record even if the form of the unit present in the text is not an exact match to the recorded

²⁵ This refers to cases where languages have gender- and number-inflected forms.

entry. Fuzzy term recognition can be manually set up if units are recorded in glossaries with a symbol such as an asterisk or a vertical bar indicating the stem of the word (e.g. *relig** for *religion, religious, religiosity*, etc.) and the TEnT is able to identify any word in the text that matches the stem pattern²⁶. Fuzzy term recognition can be automatic if the TEnT offers the option of setting a fuzzy factor (a minimum matching threshold) that units must satisfy for term records to be retrieved²⁷.

Finally, TEnTs may also include a stemming feature to increase term recognition²⁸. In this case, linguistic rules or statistical algorithms are applied to units in the sentence to be translated and recorded units in order to identify their stems. This strategy allows a record for a term to be identified when any derived or inflected form of the term appears in a text. For example, a record for *invest* would be identified based on an occurrence of *investing, invested, investment, investor, investors*, etc.

The advantage of fuzzy term recognition or term stemming is increased termbase matches. However, such features have their drawbacks. On the one hand, although both strategies will retrieve related term records, one must keep in mind that the equivalent found will not be in the exact form found in the text to translate. Therefore, the translator will have to modify and adapt the retrieved term to match the form required in the translation. While these results are useful in cases where the translator does not know the equivalent of a unit, the required modifications will eliminate the benefits of the one-click insertion. Moreover, in the particular case of fuzzy matching, the translator is forced to find a compromise between noise – where a low fuzzy factor yields undesired results – and silence – where a high fuzzy factor excludes potentially pertinent matches that are less similar to the original form.

²⁶ MemoQ and Wordfast are examples of TEnTs that offer this feature.

²⁷ Wordfast, Déjà Vu and Transit are examples of TEnTs that offer this feature.

²⁸ MultiTrans is an example of a TEnT that offers this feature in server termbases.

For the reasons outlined above, in translation-oriented integrated termbases it will be more productive to enter units not solely in their base forms but instead or also in their most frequently occurring forms. Kenny (1999) and Bowker (2011) report in their articles that translators using integrated termbases are indeed recording inflected forms.

4 Methodology

This chapter presents a more detailed description of the methodology selected to collect data for this project, in particular, the type of survey we worked with, and the approach used to analyze the collected data.

4.1 Selecting a data collection approach

The preliminary hypotheses formulated in the second step of this research had to be tested in order to assess their validity.

One option would have been to test these hypotheses in an experimental environment to evaluate which ones contribute to facilitating the translation process as represented by a model of this process based on a case study. However, such an approach would have left out a key fact about this research question: there are currently large numbers of translators using TEnTs and their integrated termbases. These translators have risen to the challenge of designing and exploiting an integrated terminology database and have done so in a great variety of ways.

Given the hope that the guidelines resulting from this research will help real future users of TEnTs, we felt that it would be invaluable to take into account the lessons learned through a wider range of translators' own experiences with integrated termbases.

Therefore we decided to assess the validity of the preliminary hypotheses using two surveys: in step three of this research we obtained a snapshot of how integrated termbases are currently used in order to compare the preliminary hypotheses to established practices, and in step five we tested only the hypotheses that, in step four, could not be confirmed or overtly refuted by users' current terminology management practices within TEnTs.

There are multiple methods of obtaining an overview of the practices of a group: case studies, experiments, analysis of discussion boards, mailing lists, interviews, surveys, etc. Below are the reasons for which we opted for a survey as our data collection approach.

Case studies and experiments are research methods that can help to validate hypotheses, but none of these will be adequate for observing trends and general practices, which was the main goal of the first survey. A case study can focus only on a very limited number of subjects (sometimes only one) because the goal of the research is to study the practices of each subject in depth in a very specific setting. This method can, for example, prove very helpful to observe whether the implementation of a series of guidelines has a positive or negative effect on an individual or group work process. An experiment can involve more subjects than a case study, but in this case engaging enough individuals to obtain a representative sample would have been very costly and the results very difficult to analyze. Experiments or case studies are avenues that this research may follow at a later stage.

Discussion boards or mailing lists of TEnT user groups provide information on users' opinions and practices. These sources have invaluable advantages (McBride, 2009, p. 69): the fact that discussion boards and mailing lists are available online allow researchers to reach a broad range of participants without geographical limitations; the pool of participants is filtered by the nature of the discussion board or mailing list (i.e. participants in TEnT usage discussion boards and translation mailing lists will be users of TEnTs or translators, respectively); data is readily available and the researcher knows the level of participation in a specific discussion topic or threads (as opposed to surveys, where response rates are only known after the fact); and finally, contributions to discussion boards and mailing lists are spontaneous, motivated only by the participant's interest and not restricted by the

researcher's questions or pre-set answers.

However, discussion boards and mailing lists present certain shortcomings. First, the research on TEnT users' profiles, their perception of TEnTs and their practices on the use of the integrated termbases as well as the research aiming to validate the sub-hypotheses seek answers to very specific questions on a particular sub-domain of TEnT usage. The free and spontaneous nature of participants' contributions to discussion boards and mailing lists would have meant having no specific control on the actual topics discussed and relying on luck to find comments on every aspect of interest for this research. Second, it is very difficult to extract commonalities and general practices from free contributions. Assuming that enough data on all aspects of the topic was available, devising a system to classify and analyze it would have posed a serious challenge. Finally, contributors to discussion boards and mailing lists identify themselves often only with a username (and in rarer cases also include their nationality or current location and years of experience). The lack of information on the participants' background would have made it impossible to compare results from different groups of users (e.g. according to years of experience, work setting, etc.)

Therefore, we opted for the survey as data collection method to investigate current trends in the management of integrated termbases.

In step five of this research, we also opted for a survey as the method to test the acceptability of the proposed preliminary hypotheses.

The obvious alternative in this case would have been a situational experiment in which users would have been provided with access to a TEnT tool, texts, different termbase resources and tasks to carry out. While such an experiment would have certainly been very interesting, it would have presented a number of design challenges that would have been difficult to overcome, including:

- Finding a large enough group of participants able and willing to come to the University of Ottawa or setting up an environment for participants to connect via the web to a TEnT hosted remotely.
- Ensuring that all participants have similar, and an advanced, level of familiarity and comfort with a particular TEnT tool, which may or not be the one they usually use.
- Ensuring that all participants have a similar experience in translation, level of source and target language competency and subject-field knowledge.
- Finding texts and examples with a similar level of difficulty for all participants.

Satisfying all the above criteria would be a very difficult task, and the results of such experiment can be very easily affected by an imbalance on any of these points. Therefore, we opted once again for using a survey as data collection method.

The following section will present the different types of surveys available and their advantages and drawbacks.

4.2 Selecting a type of survey

There are basically two main types of surveys: interviewer-administered surveys (face-to-face or over the telephone) and self-administered surveys (by mail or online) (Newman and McNeil, 1998, p. 25; Guppy and Gray, 2008, p. 108).

Interviewer-administered surveys are generally better received by respondents than self-administered surveys due to the human interaction (Guppy and Gray, 2008, p. 144). Moreover, when a survey is administered by an interviewer either in person or over the phone, the interviewer can encourage respondents to participate, clarify questions, probe the respondent to further elaborate responses and adapt his or her approach to each respondent

in order to obtain more complete results (Guppy and Gray, 2008, p. 108) or find out the respondent's reasoning behind an answer (Newman and MacNeil, 1998, p. 28). The added value of the interviewer participating directly in the survey represents at the same time a methodological challenge as participants may react differently when faced with different interviewers (Guppy and Gray, 2008, p. 144) and interviewers' probing or their adaptation of questions may render the results obtained very difficult to compare (Newman and McNeil, 1998, p. 28).

For the purpose of this research, an interview-administered survey posed many implementation obstacles. The cost of administering such a survey (facility rental, interviewer hiring, interviewer training, transportation, etc.) is very significant (Newman and McNeil, 1998, p. 28; Guppy and Gray, 2008, p. 152), and obtaining the personal contact information for our specific sample (translators using a TEnT) would have been complicated as, although there are translator directories, these rarely list the tools that individuals use. Most importantly, opting for such a method would have both required an impractical investment of time and limited the geographic scope of the study, as organizing such a study on an international scale would have been well beyond the means available for this project. Finally, interview-administered surveys do not guarantee a respondent's anonymity as interviews take place face-to-face.

For these reasons, we opted for a self-administered survey. This type of survey is formally very demanding as the structure of the questionnaire and the order and formulation of the questions must be extremely clear, given that the respondent will not be able to ask for any clarification (Newman and McNeil, 1998, p. 25; Bourque and Fielder, 2003, p. 7). Nevertheless, self-administered surveys require very little overhead compared to interviewer-administered surveys (Guppy and Gray, 2008, p. 152; Bourque and Fielder, 2003, p. 9) and

they have the advantage of being able to reach larger numbers of people and cover vast geographical areas (Bourque and Fielder, 2003, p. 10).

Once we decided to carry out a self-administered survey, the question remained whether to opt for a mail or an online one. Mail surveys require access to a list of potential respondents' addresses (Guppy and Gray, 2008, p. 150; Bourque and Fielder, 2003, p. 15). Although there are translator directories in multiple countries, we know of no directories focusing on TEnT users. Using general translator directories may have involved contacting a large number of people who would not have met the criteria to answer the survey. A more resource-efficient strategy was required. In addition, mail surveys have a very slow turnaround as all documentation is sent and returned by traditional post (Guppy and Gray, 2008, p. 152; Bourque and Fielder, 2003, p. 24). It is also costly in terms of stationary supplies and postage. Finally, to facilitate analysis, the data from the hard copies would eventually need to be transcribed into an electronic database.

In the end, we opted for an online survey for both parts of this research that required gathering data from the community of users.

Online surveys may pose sample control issues as anyone can answer the questionnaire, or conversely it may be difficult to ensure general access to it (Guppy and Gray, 2008, p. 150; Bourque and Fielder, 2003, p. 23). In this particular case, TEnT-user forums and individual contacts exceeded the initially planned size of the sample but met the target audience requirements: active translators using a TEnT. In order to compensate for the lack of control over who answered the questionnaire, we included a series of requirement questions and respondent profile questions that allowed us to become more familiar with the final respondents and to filter results accordingly. Online survey turnaround is much quicker than mail surveys as sending and receiving times are a matter of seconds (Guppy and Gray,

2008 p. 153), and the cost is more reasonable for both distribution and conversion of data into an analysis-friendly format. Moreover, since we targeted users of electronic tools, it seemed logical that such users would have access to computer resources which would allow them to participate in an online survey, and that they would be comfortable with an online format.

Although online surveys can be considered laborious to implement if one creates a survey-specific website and database (Bourque and Fielder, 2003, p. 12), certain survey hosting websites can facilitate this task. We used the services of SurveyMonkey²⁹ as the platform to distribute the questionnaires. The website simplified the implementation task enormously as it not only provided a wide variety of question formats (e.g. open-ended, close-ended, exclusive, non-exclusive, rating scales) but once the survey was designed, it also allowed online distribution with just a few clicks. SurveyMonkey is a great support tool not only for the creation of a survey but also for its analysis, as the website automatically tallies all answers and offers data analysis features such as filters, cross-tabbing and report creation. Finally, with the use of an online survey, participants' anonymity was fully preserved as the survey did not track any personal information and SurveyMonkey offered the option not to record respondents' IP addresses. This feature is important in order to guarantee respondents' anonymity.

Given the formal demands of self-administered questionnaires, special attention was given to the wording of the questions in order to ensure that they were clear, brief and unbiased. Whenever possible, questions were designed as closed-ended in order to obtain standardized answers that allowed for easier analysis of the results. In cases where an

²⁹ For more information on SurveyMonkey, visit <http://www.surveymonkey.com>. Although SurveyMonkey offers a freely available limited survey design and management option, we were able to use their more extensive professional version tools.

exhaustive list of options could not be provided, an “Other” option was provided so that respondents could specify their answers.

4.3 Selecting a sampling design

Distributing the survey to all TEnT users (the entire population of potential respondents) was not an option. Firstly, we did not have the means to conduct such exhaustive research. Secondly, we did not have the means to identify all TEnT users, even if we had limited the geographic scope of the survey to a single country or region, as no registry of TEnT users is currently available. Therefore, we had to work only with a sub-group of users (a sample).

Sampling is an essential step in the design of a survey. The sample must represent the actual population (in this case, the community of TEnT users) if we want the statistics to be accurate (Aday, 1996, p. 115; Balch, 2010, p. 71; Beidernikl and Kerschbaumer, 2007, p. 91). For a sample to be representative, any sub-group bias (i.e., over-representation of any individual sub-group or respondents) should be avoided.

Sampling designs can be referred to as *probability* or *nonprobability* designs. Probability designs are those in which all members of the population have equal chances to have access to the survey and their participation is purely decided by chance (random selection), while nonprobability designs are those in which human judgement intervenes in some way in the selection of the sample (Aday, 1996, p. 115; Trochim, 2006).

Given the impossibility of determining the total population, we were forced to opt for a nonprobability sampling design.

4.4 Analyzing survey data

The survey responses were collected in each case over a one-month period. The *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* was open from February 15 to March 15, 2009, and the *Integrated Termbases Optimization Survey* was open from July 4 to August 4, 2011.

For both surveys, only responses of participants who completed the survey in full and who answered the mandatory screening questions positively (i.e. they consented to participate and confirmed they have a good understanding of English, that they use a TEnT, and, in the case of the second survey, that they are adults) were taken into account.

Participants who completed the survey in full may have skipped certain questions by their own choice, as only the screening questions were mandatory, or by means of skip logic, on the basis of answers provided to certain questions. A survey completed in full is one where the participant has gone through all questions and has submitted his/her results at the end.

In our analysis, we give the number of respondents who skipped the question in cases where not all respondents who fully completed the survey answered a specific question.

Most questions in both surveys were multiple-choice questions, allowing either one or multiple selections. For these questions, the analysis presents the results from the most popular to the least popular option.

The other types of question in the first survey are those where participants must either classify items into a group of categories provided (e.g. potential record fields that can be mandatory, optional or not included) or rate items on a scale (normally of 1 to 5, i.e. where one is least important and 5 is most important).

In the quantitative analysis of these questions, we considered that the clearest and most informative results on the basis of which to evaluate hypotheses were those for questions in which more than 50% of respondents chose the same option (category or score), indicating widely shared opinions that provide strong evidence for the existence of common and/or best practices in the field. These were thus the primary focus for the next steps of evaluating and revising the hypotheses. Presented after these clear results were the more uncertain results, for which different options were selected by similar numbers of respondents. These cases were considered to be inconclusive, reflecting a lack of widespread agreement on best practices. This indicated a need for further investigation.

In terms of the participant profile used for the analysis, both surveys collected a similar number of responses from participants working freelance or in-house and translator was the most common profession. In both surveys, we first presented the results obtained for the total number of participants (i.e., all work environments and professions combined).

Once this initial data analysis was completed, we carried out the same data analysis by sub-groups of users. We believed that the different working conditions, needs and backgrounds of each group might have an impact on how they decided to manage terminology within a TEnT, and, thus it seemed advisable to examine the results from respondents working as freelancers vs. working in-house and respondents in different professions (translator, terminologist, manager, etc.) for both surveys. For the second survey we also analyzed the results based on whether or not the participant had received formal education in terminology. This analysis by sub-group was performed by means of filtering the answers of participants, e.g. isolating the answers of all participants who had identified themselves as freelancers.

The sub-group analysis was carried out for both surveys and for all questions. However, the results of this more targeted analysis were only included when the answers obtained differed from the answers of the whole sample.

For certain questions, we took the analysis one step further and studied the answers of groups of respondents based on the answer provided to a previous question. To achieve this type of analysis we used a cross-tabbing technique that can also be combined with filters. This combination of data allowed us to investigate more complex question such as which type of TEnT training provider covered most frequently the TMS component in their courses. We obtained the answer to this question by cross-tabbing the question that asked participants who had provided the TEnT training they had received with the question asking them whether the training received had covered the use of the TMS component. The results of this cross-tabbing are illustrated in Figure 6.

Given that we used a nonprobability sampling design, we cannot calculate the range of error or the level of confidence that the sample is a true representation of the population (Aday, 1996, p. 116; Bethlehem and Biffagnandi, 2012, p. 445). For example, the response rate is the ratio of total completed responses to total number of invitations (AAPOR, 2011, p. 5). We cannot calculate the response rate of the surveys for the simple reason that we cannot know how many individuals received the invitation, as we do not know how many people read the TEnT-user forum post, newsletter, blog or how many people re-sent the invitation to acquaintances and friends (Aday, 1996, p. 116; AAPOR, 2011, p.38; Schonlau, Fricker and Elliott, 2002, p. 106; Ye, 2011, p. 84).

Not being able to provide probability statistics means that no inferences can be drawn about the whole population based on our surveys. However, this does not mean that the data gathered has no value. Surveys using nonprobability sampling can be very useful to

carry out research on groups that are difficult to reach (Aday, 1996, p. 116; Trochim, 2006) and to draw hypotheses that can be later explored through probability-based sample surveys or other testing methods (Aday, 1996, p. 116). In our case, we can argue that TEnT users are a group difficult to reach because there is no national (and certainly no international) registry of such users. We could have carried out the survey among users of a specific TEnT or among members of a professional association. However, such samples would necessarily have been biased as they would not represent the whole population of TEnT users, focusing rather on the users of a single TEnT or on a group of translators who are likely to have a dominant language combination.

5 Terminology management practices among integrated termbase users

The preliminary hypotheses presented above were based on the logic of the functioning of TEnTs, existing literature on the topic and our own experience. Before we could validate or reject these preliminary hypotheses, there was another important source of information to be consulted. There are currently large numbers of translators using TEnTs and their integrated termbases. These translators have risen to the challenge of designing and exploiting integrated termbases and have done so in a great variety of ways.

Given the hope that the guidelines resulting from this research will help real future users of TEnTs, it is invaluable to take into account the lessons learned through translators' own experiences with integrated termbases.

Therefore, the next phase of this research reached out to integrated termbase users by means of an online survey. The reasons why we opted for an online survey as the method to collect user data have been outlined in section 4.1.

The main goal of this survey was to collect information about the current practices that users follow when they manage terminology within integrated termbases in order to evaluate and improve the preliminary hypotheses of this project.

5.1 Survey introduction: Use of terminology management systems integrated with translation environment tools

The survey collected information in four distinct areas: a) profile of TEnT users, b) organizational approaches to terminology management within TEnTs, c) content stored within TMSs, and d) information-recording strategies.

Firstly, it was useful to create an overview of TEnT users, providing details about aspects such as their role within the translation industry, geographical location, language combination(s), size of their organization, number of years' experience working within the language industry, level of familiarity with TEnTs, type of training received on the subject and tasks carried out with TEnTs. This type of overview provided details about who the survey participants were and what type of background they had. Such details were used during the analysis of the results to contextualize these results and to carry out comparisons within groups of respondents, where applicable.

Secondly, the survey explored the level and types of organizational approaches employed for managing terminology within TEnTs. Specifically, this section of the survey sought to learn whether users had created guidelines on how to manage terminology (and if so, who had participated in designing such guidelines, and what sources had been consulted to draw up such guidelines), how many termbases they managed, whether these were shared with different types of end users, and who constituted the main user group. The results of this section provided insight into the perceptions of the relevance of terminology management among participants, information about which (if any) management strategy was in place, and information about what resources and references users turn to for guidance with terminology management issues.

Thirdly, the survey asked users to provide details about what types of information they record in their TMS, and why. For example, the survey aims to find out whether respondents record only units of lexical meaning (e.g. nouns, verbs, adjectives) or whether they also create records for phraseology, common sentences or URLs and why they choose to record a unit (e.g. because it is a key concept in a specialized field, because it is a problematic unit, because it is a frequent unit, etc.).

As translators do not necessarily have all the same needs and objectives as terminologists, it was interesting to compare the results of this third section to the types of terminological units favoured in terminography when glossaries, dictionaries and term banks are compiled for reference purposes.

Finally, the survey enquired about the information-recording strategies of TEnT users. The questions in this section focused on what types of information are added to each record, which information is mandatory or optional, whether the tool's term matching techniques are taken into account to determine how information is recorded, and which (if any) information retrieval optimization strategies are used. Given that very few specific guidelines are available with regard to translation-oriented terminology recording practices within TEnTs, this research provided a description of what approaches users are applying. This snapshot of what types of information users record per record and how they do so, paired with the previous section focusing on what units are recorded and why, provided an idea of the common practices among translators, and this will in turn be used to refine the preliminary hypotheses described in section 3.

In conclusion, the first part of the survey sought to establish the nature of the survey participants and their perceptions regarding the value of terminology management, while the second described their terminological needs within a TEnT and the recording strategies that they currently use to address them. The full survey questionnaire can be found in Appendix C. In regard to the overall objective of this research project, this survey complemented the description of terminology management presence within the translation industry and management models and strategies by providing an up-to-date account of the perception and use of the terminology features of TEnTs by the members of the industry. The resulting overview of the current terminology management practices among TEnT users

allowed us to assess the preliminary hypotheses against current usage, in order to determine whether or not the practices suggested in the preliminary hypotheses are already in use within the user community. Based on this assessment, the preliminary hypotheses were later refined.

5.2 Target respondents and criteria for participation

In order to obtain a higher number of participants, we placed no limitations on the target audience in terms of geographic location, years of experience with a TEnT, specific TEnT tools, type of occupation within the language industry or language combination. Instead, we adopted a fairly broad definition for the target audience.

As described in section 0, the only requirements that potential survey respondents had to meet were the following:

- consent to participate in the survey;
- have a working knowledge of English;
- be users of one or more TEnT tools.

To ensure that participants fell within the desired target respondents group, the survey opened with three mandatory questions addressing the three requirements above. Potential survey respondents who did not meet these criteria were thanked for their interest, but they were not permitted to proceed to the actual survey questions. They were, however, invited to leave any comments that they thought were pertinent. Additional details about the design of the survey, with regard to both form and content, can be found in the next section.

5.3 Survey design

This section will discuss how the survey was designed and created. The first sub-section will focus on the survey's actual form (number of questions, types of questions and logic), while the second sub-section will focus on the content selection and actual writing of the survey.

5.3.1 Form

The survey is divided into four parts which parallel the four objectives described above: respondents' background, terminology management planning, content selection and information-recording strategies. The survey contained a total of 69 questions and it could be completed within 5 to 30 minutes. The number of questions per section presented the following distribution:

Section	Question #	Total
Requirements	1 – 3	3
Respondent's Profile	4 – 20	17
Terminology Management Planning	21 – 40, 57 – 65	29
Content Selection	41 – 45	5
Recording Strategy	46 – 56, 66-67	13
Closing	68 – 69	2
Total		69

Table 3 Breakdown of Survey Questions

Of the 69 questions, only three were mandatory. These questions correspond to the three requirements for potential participants introduced in section 5.2. Given that the survey was distributed online, there was no control over the actual sample. Introducing these three opening mandatory questions was the only viable option to enforce the survey requirements³⁰.

³⁰ Given that the survey was anonymous, this was an honour-based exercise that relied on the premise that

This survey was created using the survey tool SurveyMonkey, as described in section 4.2. The most popular question format used in this survey is the multiple-choice question allowing a single answer (40), followed by multiple-choice questions allowing multiple answers (16), text-box questions (6), matrix-of-choices questions (5) and drop-down list questions (2).

The lack of control over the sample and the survey goal of including all TEnT users who record terminology, regardless of their work setting and practices, meant that the survey had to cover as many scenarios as possible. This resulted in a longer survey (69 questions in total), but thanks to contingency questions (skip-logic rules in SurveyMonkey's terminology), respondents were able to by-pass sections that did not apply to them. For example, if a respondent indicated that (s)he or her/his organization did not have terminology management guidelines, the respondent would move on to the next section, leaving aside all questions regarding the implementation and scope of such guidelines.

5.3.2 Content

Special attention was paid to the wording of questions to avoid any leading formulations. Multiple-choice questions and matrix questions are leading in essence, as a list of options is offered. To keep their leading effect to a bare minimum, an effort was made to provide exhaustive lists of options. When possible, options were automatically randomized by SurveyMonkey. If an exhaustive list of options could not be guaranteed, an "Other" option with a text-box field was made available. This approach aimed at minimizing the number of open-ended questions, which although they are the least leading, would pose a

respondents answered the survey truthfully. We can only hope that the survey's anonymity acted as the guarantor of the respondents' honesty. Having their identity protected, respondents had no stakes at risk and therefore should have felt fully comfortable to share their real experience.

serious challenge for analysis in a survey of this length and with this number of respondents.

The broad goal to capture a snapshot of TEnT users and their terminology management practices required, as previously mentioned, numerous questions to include as many scenarios as possible. Nevertheless, the survey could not possibly be exhaustive. In order to give respondents the opportunity to point out any gaps and express their opinions if the survey questions had not captured their particular terminology management approach, questions 67 and 69 were comments fields that invited respondents to share any thoughts on the subject that they thought they had not had the opportunity to express in the 68 previous questions.

A consequence of this approach was a lengthy survey. We took the risk of obtaining fewer responses and hoped participants did not find we were inconsiderate about their time. For the purpose of this research, we felt it was necessary to carry out an in-depth survey of terminology management practices to cover a gap in the available literature and to gather enough information to later formulate hypotheses on how to optimize terminology management best practices that reflect translators' real needs. The length of the survey did take a toll on the response rate as 56 of the 168 respondents did not complete the survey. Luckily, the distribution of the survey to a large number of forums and contacts and the great interest it generated mitigated this dropout rate.

Finally, another challenge of the exploratory nature of the survey was that we could not predict which questions would be the most useful for this research. Even though the preliminary sub-hypotheses had been formulated, the aim of the survey was to capture an overview of authentic terminology management practices within TEnTs in order to draw duly grounded hypotheses. Therefore, given that the authentic practices were not known in advance, we also did not know what area of the questioning would prove the most

enlightening. As a result, this increased the level of question granularity. In addition, we invited respondents to volunteer to participate in further surveys, in case we needed to explore in greater depth an avenue that appeared promising, albeit insufficiently detailed in this initial survey.

Once a draft version of the survey was available, its clarity, neutrality and exhaustiveness were evaluated through a pilot test using two advanced users of TEnTs and a professor in this field as test respondents. Their comments and feedback were integrated into the final version of the survey.

The survey design also took into account the University of Ottawa's Research Ethics Board (REB) guidelines in order to protect the rights and welfare of respondents. The survey was successfully submitted to the REB for approval, proof of which can be found in Appendix A.

5.4 Survey distribution

The survey was published online for the practical reasons described in section 4.2. The survey was distributed via emails and forum posts that invited recipients to visit the survey website and to distribute the message to any other eligible respondent. The invitation was sent directly to 56 personal, academic and professional acquaintances that either fulfilled the requirements to answer the survey or were likely to know people in the translation industry who would fulfill the requirements. The message was also distributed to professional associations (e.g. Canadian Association of Translation Studies³¹, Association of

³¹ <http://www.uottawa.ca/associations/act-cats/>

Translators and Interpreters of Ontario³², *Ordre des traducteurs, terminologues et interprètes agréés du Québec*³³, American Translators Association³⁴, *Associació de traductors i d'interprets de Catalunya*³⁵, *Asociación Colegial de Escritores – Sección autónoma de traductores de libros*³⁶) and online user forums and discussion lists (e.g. Déjà Vu user group³⁷, MultiTrans user group³⁸, Omega-T user group³⁹, Trados user group⁴⁰, SDLX user group⁴¹, Star Transit user group⁴², WordFast user group⁴³ and Lantra-L discussion group⁴⁴). The invitation was also picked up by Jost Zetzsche who featured it in the 134th issue of Basic Tool Kit⁴⁵, a monthly newsletter for translators on software tools for the translation industry with approximately 10,000 subscribers (Zetzsche, 2011).

5.5 Survey results

A total of 168 respondents started the survey and 112 submitted it. This means that 56 respondents quit the survey at some point without reaching the end. The high drop-out rate can possibly be attributed to the length and detail of the survey.

³² <http://www.atio.on.ca/>

³³ http://www.ottiaq.org/index_fr.php

³⁴ <http://www.atanet.org>

³⁵ <http://www.atic.cc/>

³⁶ <http://www.acett.org/>

³⁷ <http://tech.groups.yahoo.com/group/dejavu-l/>

³⁸ <http://tech.groups.yahoo.com/group/multitrans>

³⁹ <http://tech.groups.yahoo.com/group/OmegaT/>

⁴⁰ http://tech.groups.yahoo.com/group/TW_users/

⁴¹ <http://tech.groups.yahoo.com/group/sdlx/>

⁴² http://tech.groups.yahoo.com/group/transit_termstar/

⁴³ <http://tech.groups.yahoo.com/group/wordfast/>

⁴⁴ <http://www.geocities.com/Athens/7110/lantra.htm>

⁴⁵ <http://www.internationalwriters.com/toolkit/>

Of the 112 submitted responses, one indicated not having a working knowledge of English and seven were not TEnT users. Therefore, the final number of respondents who completed the survey and met its requirements totals 104. Since only the three initial questions were mandatory, and some questions were skipped as a function of answers to previous questions, the fact that 104 eligible people completed the survey does not imply that they answered each and every question

Therefore, the analysis of the results was carried out only on the 104 completed surveys from respondents who met the requirements. The response rate of each question was taken into account when analyzing the results.

The following sections will present the survey results. The discussion of the results will follow the four main sections of the survey: respondents' profile, terminology management planning, content selection and recording strategy. In each section, the survey results will be described and then these results will be compared to the existing literature. Finally, we will interpret the significance of the data and observations collected in the Respondents' Profile and Terminology Management Planning sections of this study in section 5.6 and contrast the data and observations collected in the Content Selection and Recording Strategy sections with the preliminary hypotheses in chapter 6 in order to assess which sub-hypotheses are confirmed, rejected, or need further testing.

5.5.1 Respondents' profile

In order to establish a portrait of the type of language professional who participated in the survey, respondents were asked a series of questions about their background. The analysis of this section is broken down into four main areas: professional background,

technological background, training received and perception of TEnTs.

5.5.1.1 Professional background

As mentioned above, the survey was distributed world-wide and had respondents from across the globe. Only four countries had more than 10 respondents. The country with most respondents was France with 13, followed by Canada and Spain with 12 each, and the United States with 11⁴⁶. The countries with more than 5 respondents were the United Kingdom with 8, Germany with 7, and Portugal with 6. The remaining respondents originated in very different areas of the world, including Argentina, Austria, Belgium, Brazil, the Czech Republic, Denmark, Greece, Ireland, Italy, Japan, the Netherlands, Norway, Romania, Slovenia, South Africa, Sweden, Switzerland and Ukraine.

In terms of occupation, 74% of the respondents identified themselves as translators, 7.7% as terminologists, 5.8% as company or section managers, 3.8% as project managers and 1.9% as revisers. We can conclude that the survey reached an appropriate type of audience based on the sample filtering criteria and that the 77 responses from translators are sufficient. However, it must be noted that for the other occupations the survey reached a very small sample.

If we look at the breakdown of the respondents by type of employer, 48.1% of respondents were freelancers, while 21.2% worked as part of an in-house team of 2-9 members, 13.5% worked in-house on their own, 10.46% were part of an in-house team of 10-49 members and only 1.9% belonged to an in-house team of 50+ members.

The most common source language reported was English, with 59.8% of respondents

⁴⁶ Most of the individuals contacted directly resided in Spain (my native country) and Canada (my adoptive country) and all professional associations contacted but one, the American Translators Association, belonged to one of these two countries.

translating from this language. Target language distribution was more varied, yet still dominated by English with a 27.5% presence⁴⁷. The second most frequent target language was Spanish.

As illustrated in Figure 2, respondents reported that they were mostly specialized in information technologies (43.7%) and engineering (35%). However, other areas of specialization also had a strong presence, such as marketing (20.4%), law (19.4%), finance (18.4%), health (18.4%) and pharmaceuticals (13.6%).

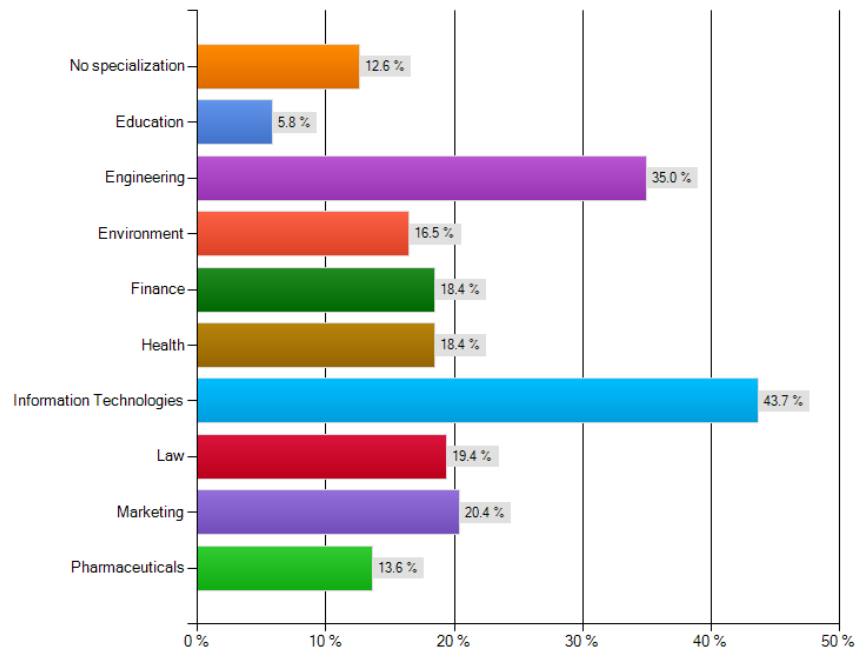


Figure 2 Respondents' Areas of Specialization

⁴⁷ Note that the survey question asked participants to select their source and target language and not to provide language pairs. This explains why English can appear as the dominant language in both categories.

As for their years of experience in their occupation, 53.9% of survey respondents had over 11 years of experience, 25% had between 6 and 10, 21.2% had fewer than 5 years of experience, while 13.5% had more than 25. Results show that 42.7% of respondents were aged 35-49 years, 30.1% were over 50, while 25.2% were between the ages of 25 and 34 years old and merely 2.2% were between 18 and 24 years old.

5.5.1.1.1. Results in context

The preponderance of translators in the respondent pool is similar to results of Lagoudaki's (2006, p. 9) survey on TMs, in which 90% of respondents were translators, the results of the UK Institute of Translation & Interpreting survey (2007, p. 4) with 85% translators, OTTIAQ's 2006 survey (p. 1) with 93.3% and SDL's terminology survey (2008, p. 1) with 91%⁴⁸.

The likely reasons for this trend are the following:

- Translators largely outnumber terminologists in the language industry. Champagne (2004a, p. 19) estimates there are 2,200 to 2,500 terminologists in Canada out of the 16,230 language industry professionals identified by Statistics Canada (2006). Therefore, only 15% of language professionals are terminologists. Very similar numbers can be observed in the results of the *Survey of the Canadian Translation Industry* (CTISC, 1999, pp. 4, 20), which estimates that in Canada there were a total of 9,135 independent and salaried translators out of the 11,790 language professionals listed by Statistics Canada in 1995. This means that in 1999, approximately 77.5% of language industry professionals were translators.

⁴⁸ *Translation Memory Surveys* by LISA (Lommel 2002 and 2004) do not make the distinction by language profession but by job category (e.g. manager, C-Level, consultant, engineering, education, GILT professional).

- Terminology is often carried out as part of the translation process by translators and not by full-time terminologists (Lommel, 2005, p. 2; Champagne, 2004b, p. 30; Jaekel, 2000, p. 163; Joscelyne, 2000, p. 91).
- This survey was not distributed to any association or forum geared exclusively to terminologists or any other language professional group.
- TEnTs were initially designed with the translator as main target user and have been marketed primarily to this sector (Lagoudaki, 2006, p. 9).

Other surveys distributed online to a broad audience seem to present a tendency to attract a high participation of freelance translators even when the survey focuses on terminology: in Lagoudaki's survey (2006, p. 9) 90% of respondents were freelancers, and in the UK Institute of Translation & Interpreting survey (eColoTrain, 2007, p. 4), the eColore survey (Wheatley, 2003, p. 3) and SDL's 2008 terminology survey, freelancers' response rates ranged between 81% and 84%. The exception is OTTIAQ's 2006 (p. 2) survey, which reports only 60% of responses from freelancers and a diminishing presence of this type of respondent over the years. In this survey 48.1% of respondents identified themselves as freelancers. Compared to the results of the above surveys, the number of respondents who worked as freelancers is substantially lower.

From these numbers one could venture to conclude that there is a higher number of freelance than salaried translators. However, the *Survey of the Canadian Translation Industry* (CTISC, 1999, p. 4) estimates that in 1999 Canada had 4,500 freelance translators, which amounted to 38% of the total population of translators established by Statistics Canada at 11,710 members. The high participation of freelancers in the above-mentioned surveys may be due to the fact that freelancers are more likely to have a higher presence online, on the networks used to distribute the surveys (e.g. forums, distribution lists, and translation

portals) or perhaps they are more likely to belong to professional associations. This may be due to the advantages of belonging to a community or because professional certification helps them to secure contracts.

As for the fields of specialization, the survey parallels the findings of Lagoudaki, who pointed out that technical fields of specialization were predominant among TEnT users. In Lagoudaki's survey, 61% of respondents specialized in technical fields (2006, p. 12). A direct comparison with the results of Lagoudaki's survey is not possible because multiple choices were allowed for this survey, while in Lagoudaki's survey, respondents were limited to a single choice. In addition, the specialization categories differ: this survey divided technical specialization into engineering and information technologies, included categories that were not listed in Lagoudaki such as health, pharmaceuticals and education, and excluded literature.

An interesting result with regard to area of specialization is the fact that 12.4% of respondents do not consider themselves to have any specific area of specialization. This result clashes with the widely accepted premise that TEnTs are tools best suited for use by translators who work in highly specialized fields where the language and text types encountered often tend to be restricted and formulaic, but would seem to support the trend that García (2006, p. 102) identified when analyzing discussions on a translators' mailing list, where non-specialized translators identified themselves as systematic TEnT-users and praised these tools for their ability to retrieve term equivalents and sub-segment matches.

The significance of these findings for this project will be discussed in section 5.6.1 below.

5.5.1.2 Technological background

Figure 3 shows that SDL Trados was the tool reported as being used (and owned) by the largest number of respondents (68%), followed by Déjà Vu (24.3%), WordFast (18.4%), Star Transit (17.5%), MultiTrans (9.7%), MemoQ (9.7%), Terminotix LogiTerm/LogiTrans (8.7%), Across (8.7%) and Omega-T (7.8%). Another interesting finding regarding TEnT use is that respondents did not report limiting themselves to working with only one TEnT. -

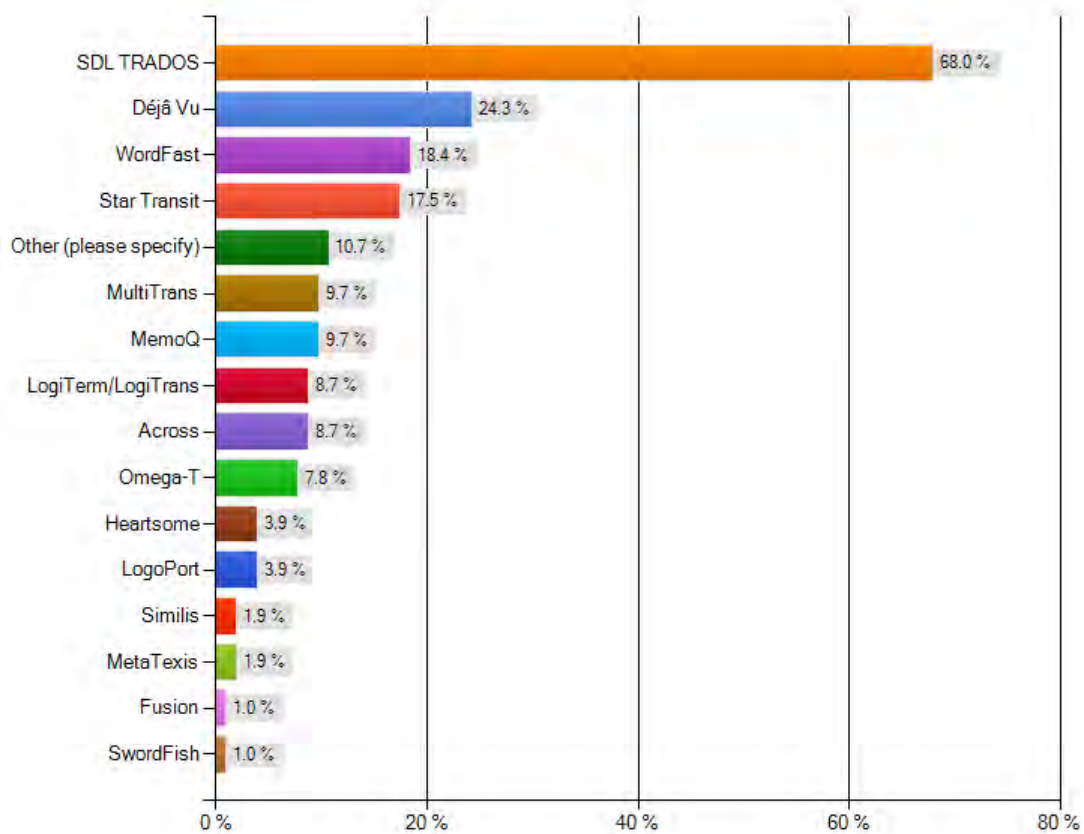


Figure 3 Distribution of the Use of Different TEnT Tools

On average respondents used 1.82 tools. If we crosstab these results by work setting, we find that freelancers own more complementary tools with an average of 2.04, while in-house translators have an average of 1.6 tools. The 2 respondents in large in-house teams of over 50 members both indicated they used a single tool. Respondents belonging to teams of 2-9 members or 10-49 members used 1.45 tools on average. In-house translators working on their own had the highest average of the in-house group at 1.6 tools per respondent, but it should be noted that this result does not exceed the overall group average and falls 0.44 behind the average number of tools owned by freelancers.

The trend of using multiple systems validates the inclusion in the survey of the next question: if you own more than one TEnT, which do you consider to be your *main* TEnT?

When respondents were asked which tool was their main TEnT, distribution did not undergo any major alterations regarding the more popular tools: SDL Trados and Déjà Vu remained the first and second most commonly-used TEnTs with a response rate of 49% for the former and 14% for the latter.

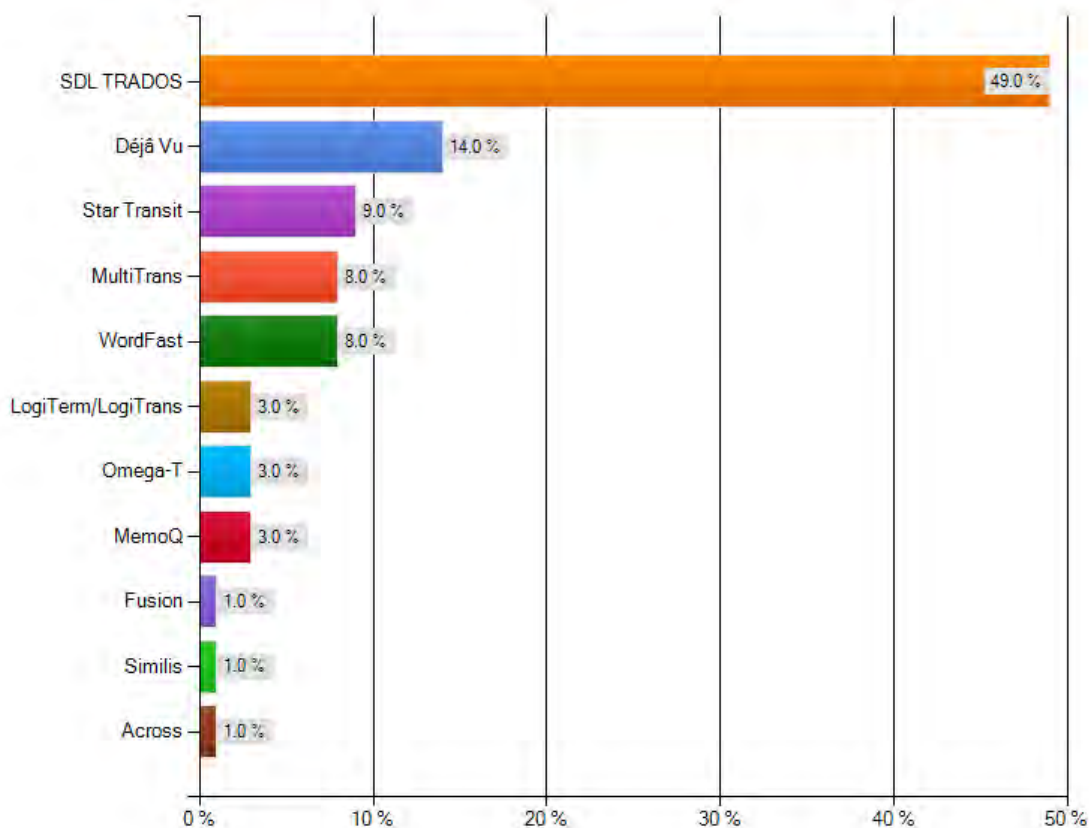


Figure 4 Distribution of Main TEnT Tools Used

The question of how respondents use their tools to translate is of even more importance to the objective of this survey. As no translation project is the same, respondents were allowed to choose more than one option to account for their own different uses of the tool. (Therefore, percentages do not add up to 100.) According to the results obtained, the vast majority (78.8%) use their TEnT to translate documents interactively. That is to say, they go through the document in a linear sequence while the tool proposes past translations of the same or a similar segment for the translator to integrate, modify or retranslate for the current target document. A sizeable portion of the sample, 29.8%, turns to another type of translation automation process: pre-translation. In this case, translators use their TEnT to

globally replace any sentence for which the tool finds a match (exact or fuzzy), and then translators edit the resulting hybrid text. A small portion of respondents, 15.4%, only use their TEnT as a reference tool to carry out manual searches. Finally, 6.7% of respondents indicated that they do not use their TEnT to translate. Respondents who do not use their TEnT for translation identified themselves as terminologists (42.8%), translators (28.6%), project managers (14.3%) or other (14.3%).

Another key factor is whether respondents were free to choose their main TEnT or whether it was imposed by their clients or employers. According to the survey results, 70.9% freely chose their TEnT, 17.5% adopted their employer's tool, and 11.7% their clients'. It could be assumed that this decision would depend on the nature of the respondents' work setting. Freelancers, in principle, are more independent, but they may be required by agencies or clients to work with a certain tool. In-house teams may already have a pre-established tool that respondents did not always have the opportunity to participate in choosing. In the case of this sample, both respondents working as freelancers and in-house teams were most often able to choose which TEnT they preferred to work with: 66% of members of in-house teams and 76% of freelancers made that choice freely. When respondents had not been able to choose their main TEnT, the choice was, not surprisingly, typically made by the employers for in-house team members (31.25%) and by clients in the case of freelancers (20%).

As far as their level of familiarity with TEnTs is concerned, respondents seem to be experienced TEnT users, especially given that this type of software only arrived on the market in the mid-to-late 1990s. As illustrated in Figure 5, it is remarkable that 19.4% of respondents have 10 or more years of experience, meaning that they adopted the use of TEnTs almost as soon as these tools appeared on the market. In general, moreover, the

respondents are not newcomers to TEnTs: 31.1% have used their TEnT for the past three to five years, 26.2% have used it for the past 6 to 9 years, 12.6% have used it for 1 to 2 years and, 10.7% for less than 1 year. Thus, a total of 45.6% of respondents have used their TEnT for more than 5 years, and 76.7% have used it for more than 3 years.

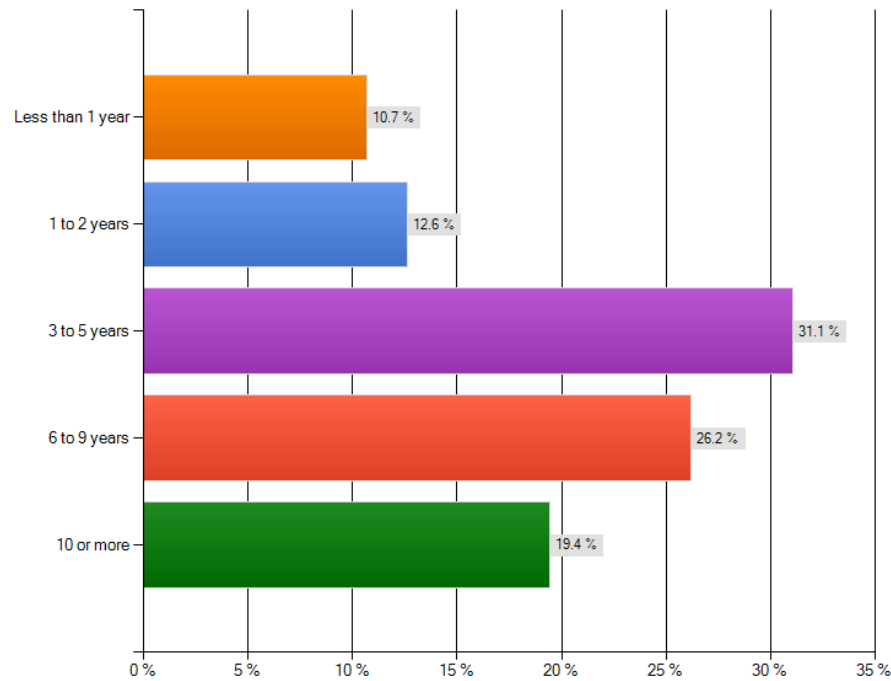


Figure 5 Experience Using TEnTs

5.5.1.2.1. Results in context

As far as TEnT brand preferences go, the results of this survey seem largely consistent with past surveys, although we must keep in mind that direct comparisons are not straightforward, given the slightly different target audiences and question formulations on the different surveys. Overall, however, Trados (now SDL Trados) remains the most widely used TEnT (see Table 4 below). Note that SDL International acquired Trados in 2005, and

thus there was a transition period overlapping with the release of SDL Trados. Déjà Vu obtained results within a percentage point in all three surveys and Star Transit ranged within a spread of 5 percentage points. WordFast seemed to have generated a surge of interest in 2006, when 29% of participants claimed to use it (Lagoudaki, p. 24) compared to only 18% in 2004 (LISA, p. 12), but it went down to 18.4% in this survey. The percentage of use of Omega-T remained relatively consistent with the results obtained by Lagoudaki in 2006 (p. 24), going up less than a full percentage point in this survey.

Survey	LISA 2004 %	Lagoudaki 2006 %	This study %
TEnT			
Trados	71	51	
SDLX	28	19	
SDL Trados		24	68
Déjà Vu	24	23	24.3
Star Transit	19	14	17.5
WordFast	18	29	18.4
Omega-T	4	7	7.8
MultiTrans	<2	<2	9.7
LogiTerm/LogiTrans	<2	<2	8.7

Table 4 TEnT Usage as Reported in LISA (2004), Lagoudaki (2006) and This Study

The two Canadian tools LogiTerm/LogiTrans and MultiTrans saw a usage increase of four times the reported usage percentages in previous surveys. It can be assumed that this surge is related to the high presence of Canadian respondents to the survey. As mentioned above, 11.5% of respondents who fulfilled the requirements and completed the survey (i.e. 12) came from Canada, where these tools are marketed more aggressively. A close-up look reveals that out of the 12 Canadian respondents, 6 own MultiTrans (accounting for 60% of the 10 total users of the tool) and 3 own Logiterm/LogiTrans (33% out of the 9 total users).

Furthermore, this survey confirmed that the majority of TEnT users do not limit themselves to a single tool. LISA's *Translation Memory survey* revealed that 57% of respondents used multiple tools, averaging 3 tools (2004, p. 11). This was also the case in Lagoudaki's TMs Survey, which showed that the average number of tools used by participants was 3.46 (Lagoudaki, 2006, p. 23). However, this survey shows a lower average of tools used at only 1.82. The results differ from those obtained by Lagoudaki, which revealed that in-house employees had access to more tools on average than freelancers (3.46 vs. 3.23 tools) (Ibid.). In this survey, the difference remains small 1.6 vs. 2.04 (i.e. .44,) but freelancers seem to be using a greater number of different TEnT tools.

Regarding the amount of experience that users have with their tools, the results of this survey were largely parallel to those of Lagoudaki (2006, p. 20), except that users who had more than 10 years of experience increased from 6% in Lagoudaki to 19.4% in this survey. This may be partially explained by the fact that this survey was conducted three years after Lagoudaki's, so a greater number of language professionals have been in a position to use tools for a longer period.

With regard to the percentage of users who were free to select their main TEnT, we see a similarity between the results obtained in this survey and those from Lagoudaki's TMs Survey (2006, p. 19): in both surveys, the percentage of respondents who were free to work with a TEnT and to choose their preferred tool came to 70%.

The significance of these findings for this project will be discussed in section 5.6.2 below.

5.5.1.3 Training received

With regard to how respondents learned to use their TEnTs, 50.5% of respondents had received formal training in this type of tool, while 49.5% were self-taught. Of those respondents who did receive training, most (52.9%) received it from their TEnT provider. The second most common source of training was industry or professional institutions (29.4%), followed by academic institutions (27.5%) and employers (25.5%). Respondents could select more than one option in case they had received training from multiple sources; results revealed that on average respondents had received training from 1.4 sources.

With particular regard to terminology management, respondents who had received training indicated that only 62.7% of the time did the training address the TMS integrated with the TEnT. Moreover, only 46.7% of the training sessions that addressed the TMS integrated with the TEnT discussed what type of terminological unit should be recorded.

Figure 6 shows that courses provided by vendors were the most likely to cover the terminology management feature (81.5%), followed by those provided by academic institutions (71.4%) and industry and professional organizations (60%). Meanwhile, according to our results, employer-organized courses were the ones that least often covered terminology-related topics (38.5%); however, only 5 respondents received training through their employers, which may not constitute a representative sample.

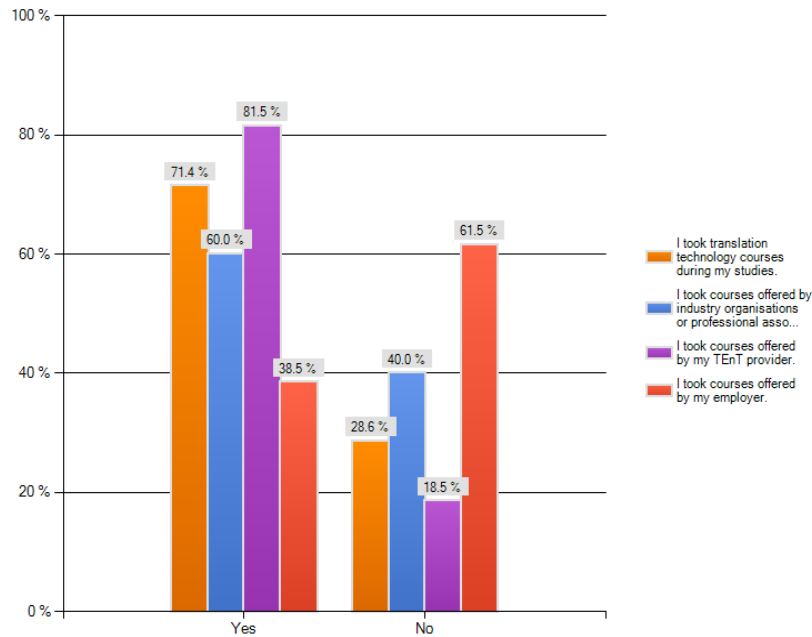


Figure 6 TMS Coverage by TEnT Training Provider

Looking into what type of training dealt not only with the TMS as a feature but also discussed what type of units to record, the only type of training that covered this aspect more than half of the time was that provided by industry and professional organizations (55.6%). When the training was provided by academic institutions and covered the TMS, in only 30% of the cases did it discuss what units should or could be recorded. In the case of courses offered by TEnT-providers that did cover the TMS, they discussed the nature of units to be recorded in only 45% of the cases, and when offered by employers, they covered this topic in only 40% of the cases. See Figure 7 for a graphic representation.

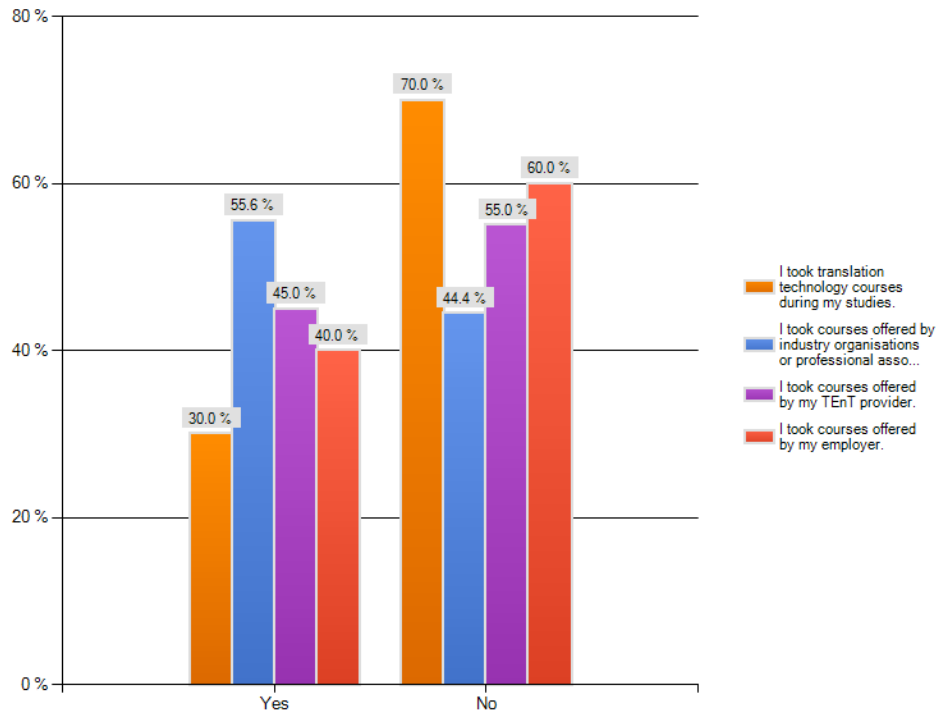


Figure 7 Content Recording Coverage by TEnT Training Provider

5.5.1.3.1. Results in context

Regarding respondents who received formal training in how to use their TEnT, results remain very much in line with the ones obtained in similar studies such as the eCoLoRe survey (in which 54% of the British respondents learned to use TEnTs on their own (Wheatley, 2003, p. 3), as did 49% of the German ones (Höcker, 2003, p. 4)) and Lagoudaki's TMs survey in 2006 (p. 19), in which 51% of respondents were self-taught. It must be noted that the above-mentioned surveys allowed only one answer to the question about how respondents had learned to use TEnTs, while our survey allowed for multiple answers.

Lagoudaki's survey also investigated the sources of the training, and it revealed that short courses and seminars were the most popular type of training while academic courses and courses from TEnT providers were the least common (2006, p. 19). The higher proportion of courses delivered by TEnT providers in our survey appears to be directly linked to the higher number of responses from in-house translators. The results for the freelance group are consistent with Lagoudaki's results: courses organized by industry or professional institutions were more common, followed by courses offered by the TEnT provider, with academic courses ranking last by a considerable margin. In our survey, courses delivered by TEnT providers seem to have become more established as they are identified as the most frequent source of training, followed by courses offered by industry or professional organizations, which tend to be either short courses or seminars. As mentioned earlier, a direct comparison of the data cannot be made as in Lagoudaki's survey respondents were asked to select only one option, whereas in the current survey respondents were allowed to indicate more than one source of training.

There are several likely reasons for the fact that only 27.5% of respondents took translation technology courses during their studies.

First, over 70% of respondents of this survey were over 35 years old, which means that—assuming they undertook their post-secondary education immediately following high school—they would have completed their studies in the late nineties or earlier, that is to say, prior to or right around the time when TEnTs started to become popular. In that case, it is highly likely that universities had not yet adopted this type of technology as part of their curriculum.

Secondly, even though translation technology is now generally included in translation curricula, it remains a challenge to deliver these courses. The eCoLoRe consortium carried out a survey of translator trainers at universities and private companies to find out the challenges they faced and to identify their level of computer skills⁴⁹. On the one hand, the survey revealed that 45.35% found teaching students computer-aided translation tools (CAT tools) to be extremely important and 25.58% found it very important (eCoLoRe, 2006, p. 22). However, only 48.8% knew how to use CAT tools and TMSs (Ibid., pp. 13, 15). The survey showed that only one third of the respondents felt fully confident to teach the use of TEnTs, 32.7% would be completely confident to teach the use of CAT tools and 25% would feel confident to teach the use of TMSs. The vast majority of respondents (84%) would themselves like to receive further training on the use of CAT tools, and 77%, on TMSs (Ibid., pp. 14-15)⁵⁰. On the other hand, the survey also points out other challenges of teaching TEnTs at the university level, such as lack of computer workstations or software tools and the fact that preparing teaching materials for TEnT courses can be more demanding due to a lack of sample materials, text types, file formats, update scenarios, guidelines on how to prepare materials, etc. (Ibid., p. 20)⁵¹.

The significance of these findings for this project will be discussed in section 5.6.3 below.

⁴⁹ The majority of respondents (74.4%) came from university professors (eCoLoRe, 2006, p. 7).

⁵⁰ In a more recent survey of 21 translation educators at the University of Ottawa, Marshman and Bowker (in press, figure 1) found similar results, with 31.9% of the respondents indicating that they were either very uncomfortable, not very comfortable, or only somewhat comfortable with technologies.

⁵¹ These findings were also supported by Marshman and Bowker (in press, figure 3).

5.5.1.4 Perception of TEnTs

In the literature review section, the importance of terminology management in translation has been discussed at length and the detrimental consequences of neglecting this task highlight its importance for any translator or translation service provider. Therefore, the survey enquired about the weight attributed to the TMS when it came time for the respondent to choose a specific TEnT.

In 52.6% of the cases the role of the TMS was considered to be important to extremely important at the time of selecting a TEnT: 17.2% of respondents considered it extremely important, 20.2% very important, 20.2% very important, and 15.2% important. See the breakdown by category in Figure 8.

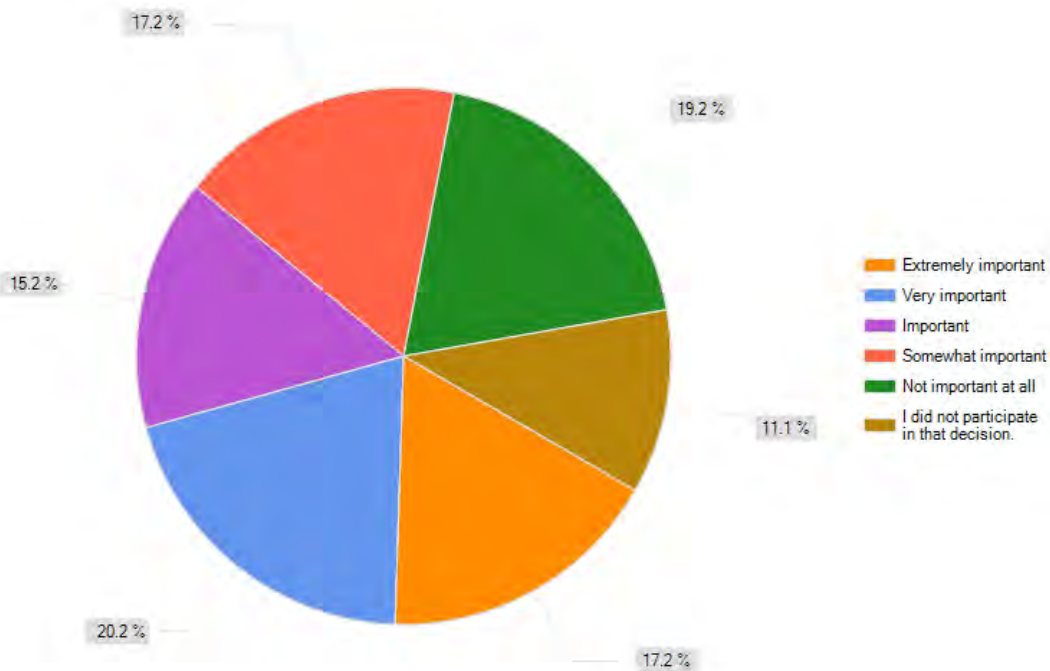


Figure 8 TMS Weight in TEnT Selection

In spite of the attention paid to this feature when respondents selected a TEnT, only 30% of respondents considered that they had mastered the advanced features of their TMSs, as illustrated in Figure 9.

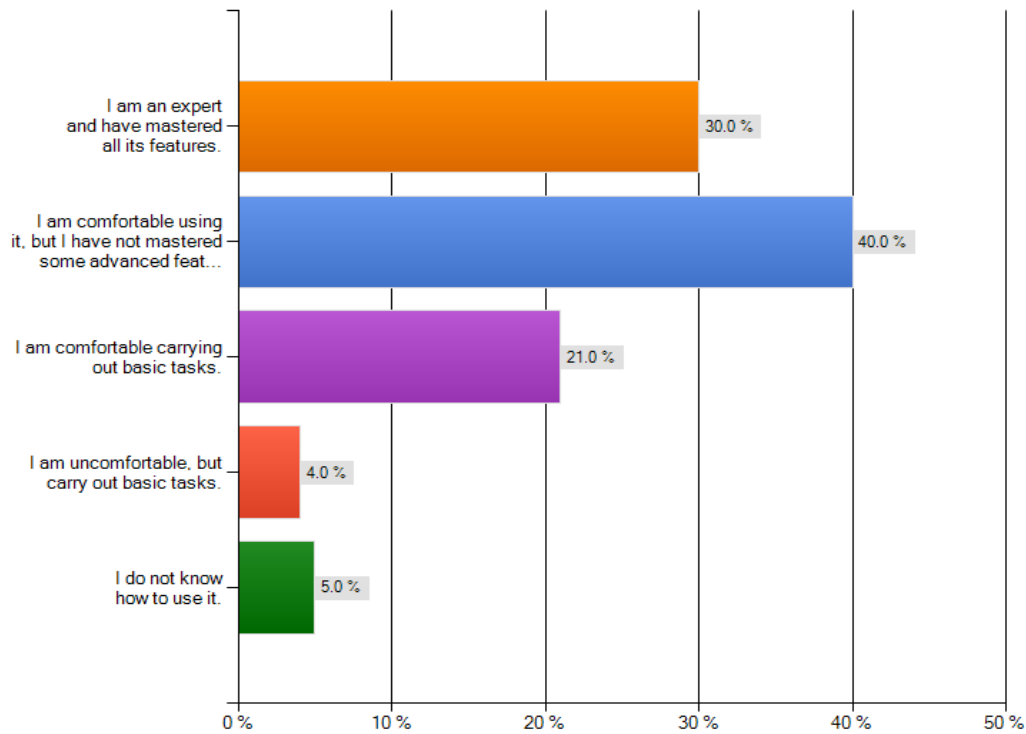


Figure 9 Level of TMS Knowledge

Although the importance of managing terminology has been clearly demonstrated within the industry, only 40% of respondents used their TMS systematically. However, this does not mean they neglect terminology since respondents spend on average 24.39% of their time per week on terminology-related tasks (i.e. an average of 9 hours and 42 minutes out of a 40-hour week). They may, however, use other tools or solutions for managing their terminology, or the time invested in terminology may be spent researching terms and their equivalents without recording their findings.

Following this question, the survey sought to discover what motivated respondents to use their TMS. A number of possible uses were provided and respondents were asked to rate them in order of priority from 1 to 5. Below are the results from most to least important. All reasons but one received the highest rating of importance (#1) by the highest number of respondents. Below they are presented in descending order according to the percentage of respondents that selected the reason as having top priority:

- To record expressions and their equivalents that required extensive terminological research. (48%)
- To create a glossary or lexicon for a specific field. (46%)
- To develop a resource that will complement the TM database and help the TEnT provide better results when translating a new document. (45%)
- To record expressions and their equivalents that I frequently look up (33%).

5.5.1.4.1. Results in context

The existing surveys and literature did not address the translator's perception of terminology management within TEnTs as directly as this survey has. Studies and surveys such as the ones carried out for the Translation Bureau's *Portrait of Terminology in Canada* (Champagne, 2004a) and *The Economic Value of Terminology: An Exploratory Study* (Champagne, 2004b), by LISA (Lommel, 2005) and by SDL (2008) inquired about whether, how, how much and why respondents managed their terminology, but their questions rarely focused on the weight of the TMS within the TEnT. However, certain parallels can be extracted to provide some context for the results obtained in this survey.

Previous surveys report on the reasons that motivate translators, translation service managers and companies to manage their terminology in broader terms. The literature review – and more specifically certain recent surveys – identify quality, consistency and productivity as reasons for managing terminology (Champagne, 2004a, p. 35; Lommel, 2005, p. 3; SDL, 2008, p. 5). The Association of Translators and Interpreters of Ontario carried out a small survey on terminology in 2008 (McInnis, 2008). Although the survey did not inquire directly why translators managed terminology, the emphasis placed by respondents on exhaustive and precise research indicates that quality is a main concern.

These general goals of terminology management are reflected in the uses of an integrated TMS that were researched in our survey.

The most frequent use selected by respondents, “to record expressions and their equivalents that required extensive terminological research,” allows the translator to avoid repeating time-consuming searches and ensures that valuable information is duly recorded. Thus, consistent terminological decisions can be made, which will ultimately improve the quality of translations.

The second most frequent use, “to create a glossary or lexicon for a specific field,” reflects a likelihood that subject-specific glossaries will help the translator to keep a record of specialized terminology, which helps to increase consistency and quality by avoiding the risk of using erroneous or inappropriate terminology.

The third most frequent reason was “to develop a resource that will complement the TM database and help the TEnT provide better results when translating a new document.” Improving the resources from which a TEnT draws information will help increase the translator’s productivity and quality because more and better results will be obtained.

In our survey only 40% of respondents used the TMS within their TEnT systematically. This result seems low, given that translators appear to be aware of the relevance of managing terminology. A possible explanation can be found in the results of other surveys in which managing terminology was also identified as a key practice. In LISA's 2005 Terminology Survey (Lommel, 2005, p. 2), 75% of respondents claimed to manage terminology systematically and 95% of translators who answered SDL's 2008 Terminology Survey claimed to spend a lot of their time on terminology-related tasks, but at the same time it was also revealed that many do not manage terminology within their TEnTs but rather prefer using spreadsheets (as was the case in 35% of the responses in LISA's 2005 Terminology Survey (Lommel, 2005, p. 4) and in 42% of the responses in SDL's 2008 Terminology Survey (p. 5)). The percentage of usage of a TMS is closer to the results obtained in ATIO's 2008 survey on terminology, which indicated that 52% of respondents had their own terminology database (McInnis, 2008). Unfortunately, McInnis does not comment on the results of the follow-up question in her survey, which inquired whether their terminology collection was stored in card files, a spreadsheet or a TMS.

Regarding the amount of time spent per week on terminology-related tasks, there does not seem to be consensus among the different surveys. Our survey's results (an average 9 h 42 min dedicated per week to terminology) were in line with the findings of Champagne (2004a, p. 33), who indicates in his *Portrait of Terminology in Canada* that translation projects require 11 h 20 min per week of terminology work. However, respondents to LISA's 2005 Terminology Survey reported investing only 3 h 48 min per week (Lommel, 2005, p. 2). The difference in these results could be due to the fact that LISA's survey targeted specifically the localization industry, where terminology is in evolution and therefore fewer resources may be invested in creating and maintaining terminology reference tools.

The significance of these findings for this project will be discussed in section 5.6.4 below.

5.5.2 Terminology management planning

The section of the survey focusing on terminology management planning was organized around the presence or absence of guidelines within an organization, along with the nature of any such guidelines. It also addressed the usage of TMSs and the ways in which termbases can be organized.

5.5.2.1 Usage

An overall objective of this research is to better understand how users manage their terminology within their TEnT environments. However, we cannot assume that respondents who use a TEnT record terminology. Unfortunately, owing to a technical problem and to the fact that some respondents declined to answer the question, the total number of respondents to the question was only 22. Of those who had the opportunity and chose to answer this question, the majority do record terminology (86.4%), while 13.6% do not.

Given that this percentage amounts to only 3 respondents, it will not be possible to consider conclusive any findings from this survey about the reasons why those respondents chose not to record terminology. Nevertheless, reasons indicated include that respondents find that other resources such as the World Wide Web, existing glossaries and dictionaries, or online corpora meet their terminological needs.

It must also be noted that not all respondents who record terminology do so within their main TEnT. While the majority of respondents (78.1%) do so, 21.9% do not. The most common reason given for respondents not using their main TEnT to manage terminology is

that they find the TMS within their main TEnT to be too complex (28.6%) or that they never learned how to use it (23.6%). Other main reasons for respondents choosing not to use their main TEnT to manage terminology are that another tool better meets their needs (19%), that their main TEnT does not meet their terminological needs (14.3%), and that a database had previously been developed on another system (9.5%).

When the main TEnT is not used to manage terminology, the tools most often used are spreadsheets and word processors (both 22.7%), in-house TMSs (18.2%), general database tools such as MS Access (13.6%), and other TEnTs or stand-alone TMSs (both 9.1%).

Respondents were asked what would encourage them to record terminology within their TEnTs. However, either due to a technical problem or because this question came towards the end of a somewhat long survey, the question received only 11 responses. Of the 11 respondents, 7 consider that access to more or different kinds of documentation would help and some specifically point to online help and training materials. Receiving more training is viewed as a way to get started on recording terminology by 6 out of the 11 respondents, and some specifically request “advice on rules for setting up a database that would be most useful”. Finally, 6 out of the 11 respondents indicated that the addition of new features to their TEnTs would also help them take the step of beginning to record terminology. To the question of what kind of new features would encourage them to begin recording terminology, two respondents suggested that a more flexible user interface, a function for sending pairs of terms from the TM to the TMS, and a system that requires less RAM would be positive improvements.

At the time of the survey, 4 out of 10 respondents who did not yet manage terminology planned to do so in the future, 4 were uncertain and 2 had no intention of starting to manage their terminology. When asked what tool they would choose to manage terminology in the future, out of 6 respondents, 3 indicated that they would select a TEnT, 2 would use a word processor and 1 would choose a spreadsheet.

5.5.2.1.1. Results in context

The literature review section revealed that companies are aware of the relevance of managing terminology and appreciate its ROI (Champagne, 2004a, p. 7). Moreover, it showed that although only 7.8% of Canadian companies overall have a terminology specialist (Champagne, 2004a, p. 17), this number rises sharply among companies within the translation industry (Lommel, 2005, p. 2; McInnis, 2008). While our survey did not ask participants whether their companies employed a terminologist, it did ask whether they used a TMS within their TEnT, i.e. whether they recorded terminology. The results revealed that 86.4% of respondents use a TMS within a TEnT.

Although limited by a small sample, the reasons for why translators do not record terminology differ greatly from the ones obtained by Lommel (2005, p. 2). Lommel's survey focused on companies within the translation and localization industry, where the reasons for not recording terminology included lack of budget, time, infrastructure or demand from clients (Ibid.). In our survey, which targeted translators and garnered a large response from freelance translators, respondents justified not recording terminology because they found their terminology needs could be met by other resources.

If these results were to be confirmed in a larger sample, it would be interesting to explore whether managers are more aware of the value of managing terminology than are translators, given that the answers in Lommel's survey seem to acknowledge the relevance of the practice and justify a lack of action on that front due to pragmatic obstacles. In contrast, the answers obtained in our survey indicate that respondents do not perceive losses incurred by having to repeat their searches.

Our survey, like Lommel's, showed that some respondents do record terminology but choose to do so with a tool other than their main TEnT. Compared to Lommel's results, a smaller number of respondents use spreadsheets or word processors. In our survey such respondents accounted for 5% of the total, while in Lommel's they made up 35% (2005, p. 4). The reasons for using these alternative tools in both cases also differ. According to Lommel's findings, spreadsheets were favoured by users of multiple TEnTs in order to facilitate portability (Ibid.), while in our survey, spreadsheets and word processors seem to be preferred for their ease of use.

Finally, even though there were a limited number of responses regarding those features that would help users record terminology within their TEnTs, the answers obtained coincide with Jost Zetzsche's criticism of the "awkward methods of entering and retrieving data" offered in most TEnTs (2006).

The significance of these findings for this project will be discussed in section 5.6.5 below.

5.5.2.2 Guidelines

Our literature review concluded that there are no specific guidelines generally available on how to organize and manage terminology within a TEnT. Therefore, an important question included on our survey was whether respondents had created any guidelines of their own and, if so, what resources they had relied on to do so. The results indicate that 53.2% of respondents do not plan what information is recorded or how, 33% have only basic guidelines regarding terminology management, and 13.8% have very specific guidelines. If we look at the results categorized by work setting (see Table 5, below), we see that freelancers are the group least likely to plan their terminology management, in-house translators working on their own tend to plan more than freelancers, and in-house teams with multiple members are the ones that plan the most.

Work setting \ Guidelines	No guidelines %	Basic Guidelines %	Very Specific Guidelines %
General	53.2	33	13.8
Freelancers	69.4	24.5	6.1
In-house translators working on their own	41.7	50	8.3
Multi-member in-house teams	33	39.4	27.3

Table 5 Presence of Guidelines in each Work Setting

With regard to the types of resources that respondents consulted to design their guidelines, the answers show that they used a wide range of available information, including vendor, industry and academic literature, academic and industry specialized courses, advice from vendor specialists and other users, their own past experience, and the reference model of existing glossaries and dictionaries. Table 6 lists the resources that the largest group of respondents selected as either very important or somewhat important. Some resources

received equal numbers of votes for different categories. The details of these cases presenting a lesser degree of consensus will be described next.

Very important resources	%	Somewhat important resources	%
Vendor documentation	45.7	Other TEnT' users' advice	50
Existing paper glossaries and dictionaries	38.2	Past experience compiling glossaries in other TEnT tools	35.3
Past experience compiling glossaries in non-TEnT tools	38.2	Recommendations by a vendor specialist	33.3
Industry organizations' documentation	31.3	Specialized courses provided by a professional association or an industry organization	25

Table 6 Use of Resources for Guideline Creation

The resources that were clearly identified by the largest group of respondents as very important are vendor documentation, existing paper glossaries and dictionaries, past experience compiling glossaries in non-TEnT tools and documentation from industry organizations. The remaining resources were generally considered somewhat important with only a few exceptions.

For academic works and academic institutions' specialized courses, the opinion was very equally divided, among the options very important, somewhat important or not important at all. The results for these resources are illustrated in Table 7 below. Finally, although a larger number of respondents considered specialized courses offered by industry and professional organizations to be somewhat important, a very similar number of respondents judged them very important and not important at all.

Resource \ Rated	Very important	Somewhat important	Not too important	Not important at all	Not applicable
Academic works	24.2%	24.2%	12.1%	24.2%	15.2%
Academic institutions' specialized courses	23.5%	23.5%	11.8%	17.6%	23.5%
Industry associations' specialized courses	21.9%	25%	21.9%	18.8%	12.5%

Table 7 Guidelines Resources with Inconsistent Results

Once it had been established which resources respondents relied on to design their guidelines, they were asked about the nature of such guidelines. First, they were asked whether their guidelines imposed any limitations on the nature of the units recorded. The results are shown in Figure 10. This question sought to verify whether traditional terminological principles such as the onomasiological approach, univocity or the preference for recording noun-based units were upheld or whether these had been set aside as certain studies have indicated (e.g. Bowker, 2011; Kenny, 1999; O'Brien, 1998). According to our results, a clear majority of 68.9% of respondents who used guidelines indicated that their guidelines did not impose any limitations regarding the nature of the unit recorded and that they were free to suggest creating a term record for any unit.

Some respondents pointed out that their guidelines did impose limitations regarding the nature of the unit. Almost a third of the respondents, 31.1%, used guidelines requiring that only concept-denoting units be recorded in accordance with the onomasiological approach, and 8.9% of respondents were not allowed to create new term records for units that were considered to be synonyms of previously recorded expressions, a limitation that is in line with terminology's univocity principle. Figure 10 shows the frequency of several guideline limitations according to the responses collected in this survey.

Limitations regarding units having to belong to a specific part of speech, i.e. nouns, were very uncommon, with only 4.4% of respondents indicating that such criteria were part of their guidelines.

Our survey went one step further and asked respondents whether the guidelines in place took into account a series of factors regarding what information to record in their termbase. The most frequent factors specified in users' guidelines that were intended to help translators decide whether a unit should be added to a termbase were frequency (47.7%), collocations (38.6%), form variation (36.4%) and syntactic structure (25%). Note also that 27.3% of respondents indicated that none of the above factors was taken into account by their guidelines.

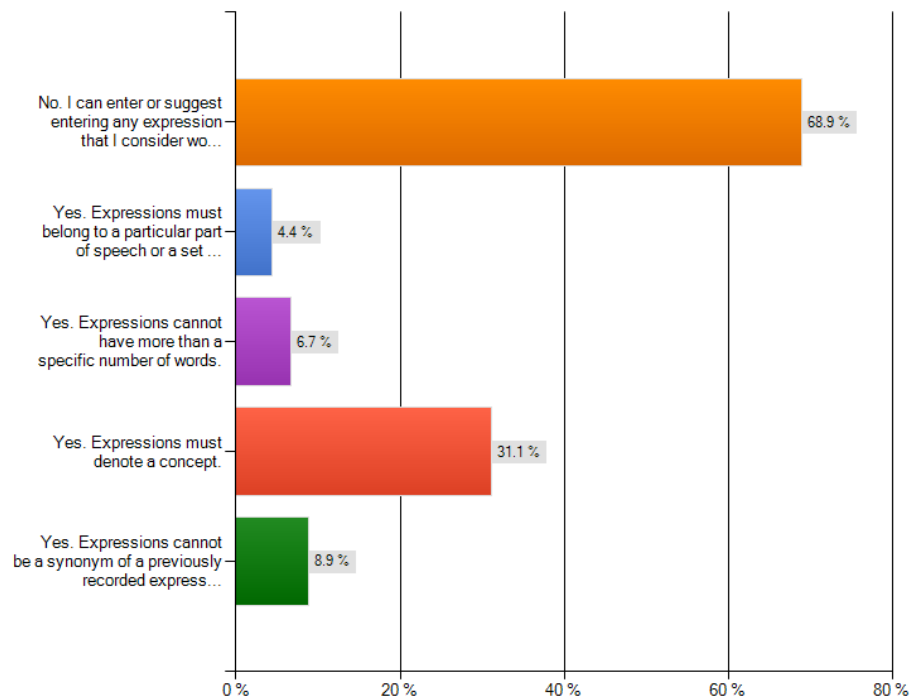


Figure 10 Guideline Limitations on Recordable Units

5.5.2.2.1. Results in context

Defining and delimiting the notion of terminological unit or term is a cornerstone of terminology theory. Numerous works have addressed this topic and strict and sophisticated principles have been established to identify terms (Cabr , 1998, p. 148; Dubuc, 2002, p. 33). In traditional terminology, an expression will be recorded only if it denotes a concept (Cabr , 1998, p. 149; Dubuc, 2002, p. 33; Pavel and Nolet, 2001, p. 18). Bowker points to the fact that translators building a termbase within a TEnT are primarily concerned with recycling the greatest amount of text from their past translations, which often leads them to include expressions that are not strictly terms as long as they occur frequently (2011, p. 14). The results of this survey support Bowker's analysis of the impact of TEnT tools on translators' term-recording strategies: only 31.1% of respondents who use guidelines include as a requirement that term units denote a concept; 68.9% can record any type of expression without limitations. Note that Bowker's analysis is further backed up by the fact that frequency is the most common factor included in guidelines with regard to which units should be recorded.

The significance of these findings for this project will be discussed in section 5.6.6 below.

5.5.2.3 Database design

A specific aspect that guidelines for managing terminology within a TEnT should address is whether to keep multiple termbases or whether to create and maintain only a single termbase. The vast majority of survey respondents prefer to maintain multiple termbases (81.6%) rather than centralize all their terminology in one termbase (18.4%).

When respondents maintain multiple termbases, these are most frequently divided by subject (71%). Other termbase organization parameters are client (51.6%), language combination (30.6%) and project (22.6%). It should be noted that these parameters are not exclusive: respondents may have termbases divided both by subject and by client or project. Respondents who maintain only one termbase seem to use multiple parameters to classify their records within that termbase. In descending order of popularity the parameters are: subject and language combination (64.3% each), project (57%), client (50%) and date (42.9%).

Our survey explored whether respondents built their termbases for personal use or to share them with others as this may influence decisions about database organization. The survey shows that 42.5% of respondents created personal termbases, but those who shared them did not fall too far behind at 32.9%. In addition, 24.7% engaged in both practices, having some personal termbases and some shared ones.

With regard to freelancers, one might have predicted that they all built only personal termbases, and although this was the case for the majority (65.8%), some freelancers built some termbases for personal use and others to be shared (18.4%), while other freelancers built exclusively shared termbases (15.8%). One out of four respondents belonging to in-house teams of 2-50+ members used shared and personal termbases. Personal termbases were more frequent among in-house translators who work on their own (50%), and interestingly, among 11% of respondents belonging to in-house teams of 2-9 members.

Shared termbases were most commonly made available to translators (95.6%), followed by editors/revisers (51.1%), project managers (42.2%), terminologists (37.8%) and clients (22.2%). In spite of the importance that those who have studied the impact of terminology management place on harmonizing terminology at the source, technical

writers/authors, employees in general and the general public were the least common users of these shared termbases, with a corresponding percentage of usage of 13.3%, 15.6% and 8.9% respectively.

5.5.2.3.1. Results in context

The fact that a majority of respondents prefer creating multiple separate termbases instead of centralizing all terminology information in one repository seems to go against the recommendations found in some of the literature. Authors such as Sofer (2006, p. 96) and Samuelsson-Brown (2004, p. 86) recommend building centralized termbases and classifying the contents at least by client if not also by domain. On Déjà Vu's user forum, keeping a single large termbase is also the approach that is presented as the most popular and is particularly recommended for translators who either do not have subject overlap among their clients or who have clients whose terminology does not conflict (ÓhAiseadha, 2001). Whenever clients in the same subject field use specific terminology that is not interchangeable, Déjà Vu users recommend either using separate termbases or classifying records by client (Bergeron, 2001).

Esselink (2000, p. 398) concurs with the opinion of our survey's respondents and, focusing on large software localization processes, he recommends creating and using different types of glossaries depending on the sources of the terminological information: software strings, terminology provided by the client or parallel documentation for the project extracted from the product website, help files or any additional documents.

Regarding freelancers' practices in sharing their termbases, the results seem surprising: 18.4% of freelancers build some shared and some personal termbases and as many as 15.8% built exclusively shared termbases. These results are unexpected as traditionally freelancers have been perceived as solitary and isolated. The arrival of new technologies and increased market demand have encouraged freelancers to network, team up and possibly create a different reality for the freelancer than the stereotype that may be commonly espoused.

ATIO's 2005 *Survey of Independent Translators* reported that 41% of respondents used Internet forums for help. Recent research has opened the debate on this metamorphosis of the freelancer's work setting. In her categorization of translation networks, McDonough (2007, p. 797) points out the popularity of "practice-oriented networks", which abound on the WWW and can have memberships of all sizes (from a dozen to hundreds of thousands), create links among their users in which they may exchange "material goods" (i.e. translations) or "non-material goods" (i.e. advice and opinions) (Ibid., p. 803).

The results of our survey seem to indicate that not only are freelancers open to helping each other find translation solutions and teaming up to undertake large projects, but they also share their linguistic resources. Further research would be necessary to determine whether freelancers are sharing their databases with fellow translators with whom they work on a project, with translators with whom they do not share a common project, or with their clients.

This finding collides with the diminishing interest in the exchange and purchase of TMs reported by LISA, when comparing the results from its 2002 *Translation Memory Survey* (Lommel, 2002, p. 16) and their results from the homologous survey carried out in 2004

(Lommel, 2004, p. 16). In 2002, 50% of respondents were interested in exchanging their TMs, while in 2004 that number decreased to 24%. LISA's 2004 *Translation Memory Survey* also revealed a larger interest in TM exchange by translation service providers and researchers than by consumers (Lommel, 2004, p. 17).

In subsequent years, this interest proves to have persisted and has actually translated into at least two for-profit initiatives to pool TM resources. On the one hand, in 2005 TM Marketplace⁵² was launched by MultiLingual Computing Inc. in partnership with International Writers' Group, targeting all users from freelancers to corporations. TM Marketplace offers a series of TMs with different levels of quality and matching ratio guarantees that can be purchased or licensed. On the other hand, in 2009, TAUS (Translation Automation User Society) launched the TAUS Data Association (TDA)⁵³, a membership-based association that centralizes data from major corporations and shares it with its members based on their membership level.

TEnT end-users and freelancers may be less open to sharing their TM databases than their terminology due in large part to the intellectual property debate, as pointed out by Lommel (2002, p. 16; 2004, p. 17). The question of who owns the copyright to the translated text has been the subject of study for years in the translation field, and consequently, this debate extends to the rights over TMs (Bowker, 2002, p. 122; Topping, 2000), which are collections of translated texts or segments. Proof of the debate and the difficulties involved in resolving it owing to its abstractness and the different intellectual property laws found in different countries is evidenced in a study carried out by Francie Gow and reported in "*You Must Remember This: The Copyright Conundrum of 'Translation Memory' Databases*" (2007, pp. 175-192). In this study Gow points out that clients can most easily trade TM databases because

⁵² <http://www.tmmarketplace.com>

⁵³ <http://www.translationautomation.com/tda/mission-a-activities.html>

they are the ones who are most likely to own full rights over both source and target texts (2007, p. 183). This is precisely the case in the exchange within the TDA, where clients contribute their own TMs. The reason that Gow suggests for the infrequent commercialization of such products by clients is that the client's main business goal is not selling such assets but rather selling other goods or services (2007, p. 188). This may be true for small clients, but as the presence of large international corporations (e.g. ABBY, Adobe, Cisco Systems, Dell, eBay, McAfee, Microsoft, and Oracle) in the TDA shows, when the demand for translation services within a company is large enough, clients may pay more attention to the linguistic assets they have in their hands.

The significance of these findings for this project will be discussed in section 5.6.7 below.

5.5.3 Content selection

Up to this point our survey focused on terminology management at a high level, asking about perception of the tools, guidelines designed, database organization, etc. The remaining part of the survey addressed the respondents' everyday tasks and decisions with regard to building their termbases. A key question in this area is what motivates respondents to decide to record a given unit.

As shown in Figure 11, the factor that most respondents (81.1%) consider very important when deciding to record a unit was that the unit should represent a key concept within a specialized field. The next factor that most respondents valued as very important when choosing to record a unit was the situation where a unit was unknown and required research (69.9%). This was followed by units being specific to a company, project or subject (58.1%), and units being merely frequent (56.9%) or likely to lead to error (e.g. homonyms,

special grammatical structure, connotations, etc.) (55.6%). Less than half of the respondents (49.3%) considered that not knowing a unit's equivalent was a very important factor when deciding to record that unit. However, 38% considered the factor somewhat important, leaving only 12.7% of respondents considering it not too important or not important at all. The factor that received the lowest number of very important ratings and a comparable number of very important and not too important answers is whether a unit is a proper noun (person, institution, document, product, etc.).

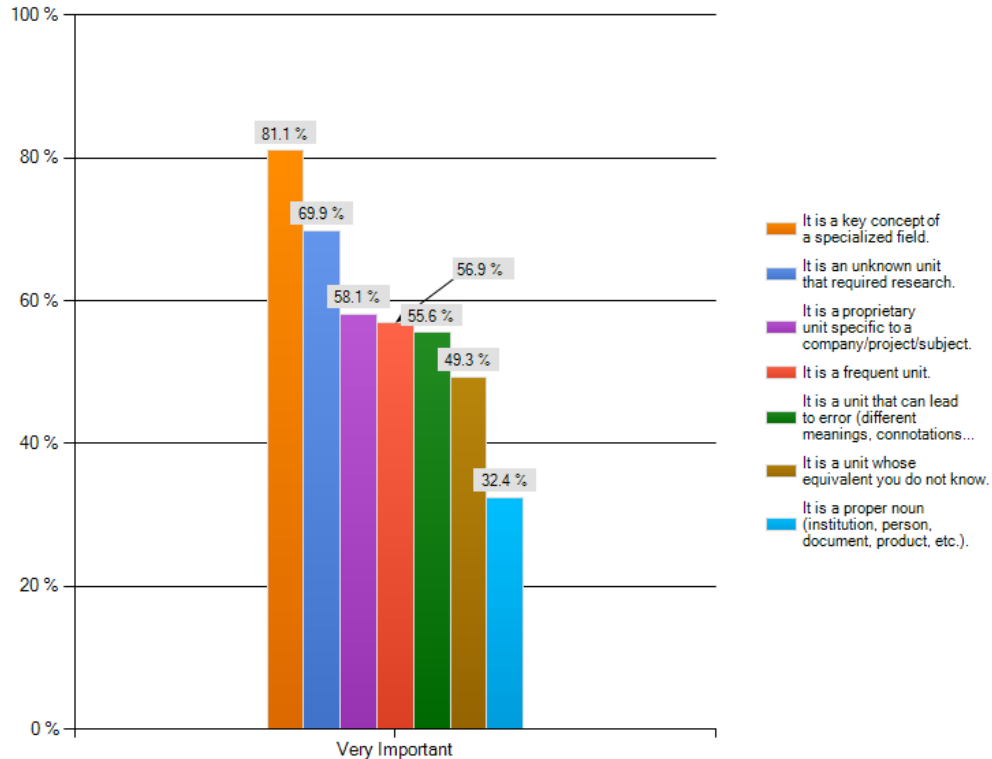


Figure 11 Comparison of Unit Selection Criteria Ranked as Very Important

This information can be complemented with knowledge about what types of units respondents select to record in their termbases. Respondents were asked to indicate all the types of units for which they currently had records, and therefore they were able to select more than one option, as shown in Figure 12. The most common unit type was the noun (93.2%), followed by the verb (82.2%), the adjective (78.1%), the phrase (e.g. article + noun, verb + preposition, adjective + noun) (72.6%) and the proper noun (63%). Slightly less than one third of the respondents also record frequent sentences (32.9%), and only 16.4% record frequent paragraphs. Few respondents record URLs (8.2%), and very few record email addresses (1.4%). None of the respondents recorded telephone or fax numbers or physical addresses on term records.

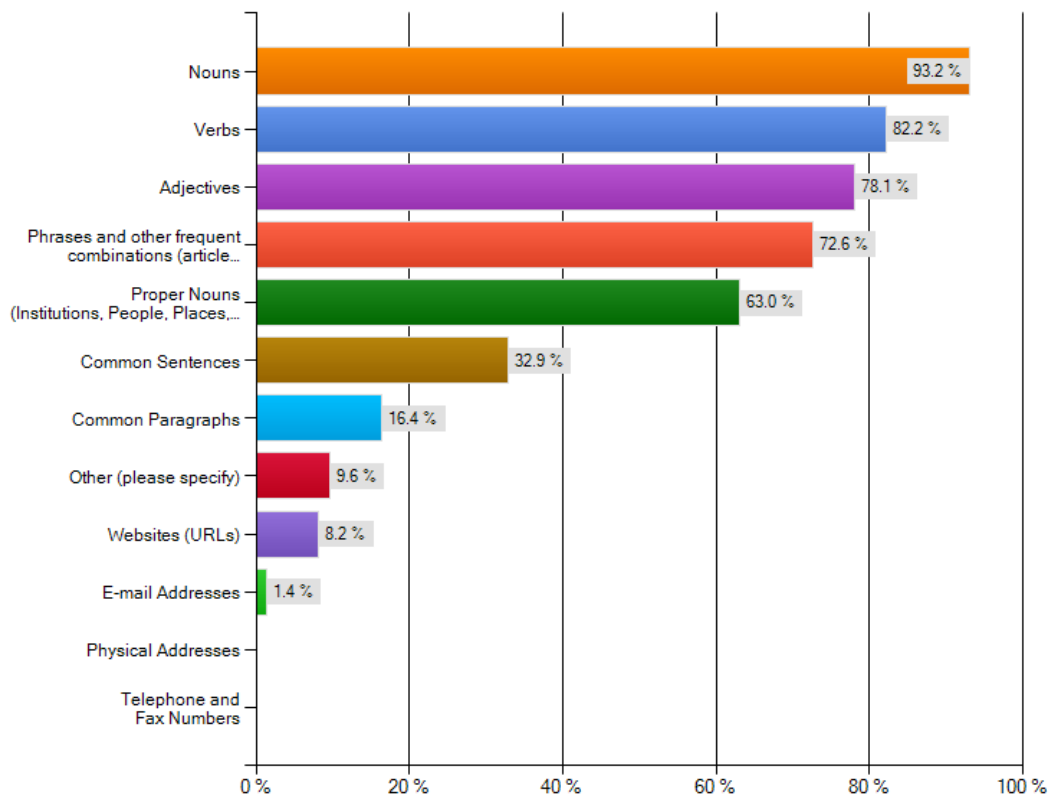


Figure 12 Unit Types Recorded

Our survey also sought to better understand the impact that the TEnT had on the creation of a termbase within it. The survey results indicate that, when selecting what terms to record, 69.4% of respondents take into the account the ability of the TEnT to retrieve and insert terminology during the translation process.

As described in detail in section 1.1.1, TEnTs compare a new document to be translated against existing TM and terminology databases, if available; results from both these sources are displayed. However, the tool chooses which match is more relevant for automatic insertion in the translation pane as its top proposal. This choice is made according to a set of rules, which vary depending on each TEnT.

In spite of the central role these rules play in deciding what information is provided to the translator, only 44% of respondents to our survey were familiar with their TEnT's resource prioritization rules and take these into account when selecting their termbase content. As many as 33.3% of respondents are not aware of such rules and 22.7%, although familiar with them, do not take them into account when deciding which units to record.

Building a termbase within a TEnT may be influenced not only by the ability of the TEnT to retrieve this information, but also by the capacity of such tools to build bi- or multilingual parallel corpora (i.e. TM databases). The survey inquired whether respondents used such resources as reference sources for their terminological research. Results indicate that 46.1% of respondents include TMs among the resources they consider when researching a unit, and 39.5% consider them to be one of their top resources, while 9.2% do not use them for terminological research and 5.3% consider them a last resort.

5.5.3.1.1. *Results in context*

Within the traditional terminology literature, it is an *a priori* requirement that a unit must denote a concept in a specialized field in order to qualify as a term and to be recorded on a term record (e.g. Cabré, 1998, p. 149; Dubuc, 2002, p. 33; Pavel and Nolet, 2001, p. 18). Although this aspect is also considered to be very important among respondents to our survey, other aspects are also considered important, such as the need for research, units being proprietary, their frequency or their potential to lead to error.

The diminished emphasis on the criterion of “termhood” may be explained in part by the fact that translators’ needs differ from terminologists’ needs. The main goal for translators is to render a text in another language, and therefore they can be expected to be especially interested in any unit that requires research or special consideration for the purpose of translation (Estopà, 2001, p. 17; Bowker, 2011, p. 22). This is reflected in the ranking of their motivation factors.

As for the type of morphological units extracted, the results obtained follow the premises established by terminology theory. Respondents create records for nouns, verbs, adjectives and proper nouns. Terminology literature often establishes nouns, verbs, adjectives and phrases as the morphological structures in which terms can appear, including complex source nouns such as initialisms and acronyms (Cabré, 1999, p. 139; Dubuc, 2002, p. 34; Pavel and Nolet, 2001, p. 19; L’Homme, 2004, p. 64). Some works also include as possible terms adverbs (L’Homme, 2004, p. 64; Translation Bureau, 2008b) or even official titles, administrative entities or working titles (Translation Bureau, 2008b).

Regardless of the units accepted as terms, most authors concur that terms are more commonly found and most easily spotted in the noun form (Cabré, 1998, p. 57; L’Homme, 2004, p. 59; Translation Bureau, 2008b). As established in the founding General Theory of

Terminology, concept-based terminology or denominational terminology only considers as terms those units that refer to a concrete or abstract object of a specialized field (Cabr , 1998, p. 168; Dubuc, 2002, p. 33). The noun is therefore the unit *par excellence* that can fulfill the requirements above, which also explains its popularity among the responses to this question.

Research on translators' terminological habits or recommendations expands the list of units considered interesting to record for translation purposes to include slogans, formulas, phone numbers (Kenny, 1999; Bowker, 2011), frequent words, phrases or sentences, software strings and transposables (i.e. units or segments that do not require translation) (Esselink, 2000, p. 400).

While 56.9% of respondents recorded units based on their frequency, only one third of respondents recorded frequent sentences (e.g. slogans, formulas) and only around 14% record frequent paragraphs. Results also show that respondents rarely recorded phone numbers, URLs or physical addresses.

Finally, terminology theory favours the use of original-language documents over translated texts as sources for terminological research (Dubuc, 2002, p. 156; Pavel and Nolet, 2001, p. 8). Bowker (2011, p. 221) points out that although using source texts is the ideal approach, translators may turn to translated documents to find equivalents either because they lack the time to carry out research in source documents or because they feel that they can trust the reliability of their translated sources. These survey results show that respondents consider their own TM databases as reliable documentation sources that often rank among their favourites, which supports Bowker's statements.

The significance of these findings for this project will be discussed in chapter 6 below.

5.5.4 Information-recording strategies

Finally, after gathering data that can provide a better idea of which units are recorded in termbases within a TEnT environment, our survey set out to explore how this information is recorded.

5.5.4.1 Designing a template

The first question on this subject inquires about what types of information fields respondents include on their term records, and whether they are considered to be mandatory or optional.

The two fields that were almost unanimously considered as mandatory among the respondents were the source (98.7%) and target (96.1%) terms. The rest of the fields were classified either as optional or as not included.

Two other fields that were included on term records by a great number of respondents were the fields for administrative information, such as client, project or date, and the domain field. Administrative information fields were considered optional by 49.3% of respondents and mandatory by 35.6%. The domain field was considered optional by 42.3% of respondents and mandatory by 38%. The remainder of respondents identified these fields as not included.

None of the other fields listed were considered mandatory by half or more than half of the respondents. Most respondents considered as optional fields such as synonyms (78.6%), definitions (69.9%), short forms (68.6%), context (67.1%), reference material (websites, documents, experts) (65.7%), cross-references and related terms (64.7%) and sub-domain (51.5%).

For all grammatical information fields (part of speech, gender, number, case) the results between optional and not included were too similar for the difference to be relevant. Regarding part of speech, gender and number, the results all fell within the 42% to 57% range in each category, and in all three cases within a span of 7.5 percentage points. A field for case form was not included in 56.7% and optional in 40.3% of respondents' answers.

Field	Mandatory	Optional	Not Included
Part of speech	10.1%	47.8%	42%
Gender	5.8%	49.3%	44.9%
Number	4.4%	44.1%	51.5%
Case	3%	40.3%	56.7%

Table 8 Presence of Grammatical Fields in Respondents' Record Templates

Other fields that produced close results between optional and not included rankings are the collocations, image and inflected forms fields. For more details on the exact percentages for these fields consult Table 9 below. A majority of respondents did not include syntactic information in their records (62.7%).

Finally, certain fields were split three ways and the results are too close to make a statement about preferred practice in these cases. This is the case for information such as the author of the term record, term source, definition source and context source.

Mandatory		Optional		Not included	
Source term	(98.7%)	Comments	(81.5%)	Syntactic information (62.7%)	
Target term	(96.1%)	Synonyms	(78.6%)		
		Definitions	(69.9%)		
		Short forms	(68.6%)		
		Context	(67.1%)		
		Reference material	(65.7%)		
		Cross references	(64.7%)		
		Sub-domain	(51.5%)		
	(35.6%)	Client, project, date	(49.3%)	Grammatical information*	
	(38%)	Domain	(42.3%)	(53.7%)	Collocations (46.3%)
				(45.7%)	Image (52.9%)
				(41.8%)	Inflected forms (56.7%)
Author	(38.6%)		(32.9%)	(28.6%)	
Term source	(32.9%)		(41.4%)	(25.7%)	
Definition source	(24.3%)		(45.7%)	(30.0%)	
Context source	(22.4%)		(43.3%)	(34.3%)	

Table 9 Record Template Model

*Part of speech, gender, number and case have been consolidated under this category given the similarity of the results

5.5.4.1.1. Results in context

Other studies have investigated what information is recorded in termbases. For instance, O'Brien (1998, p. 188) concluded that in the localization industry such databases tend to record only source and target terms given the short lifespan of terminology in this field. According to the results gathered in our survey, source and target terms would indeed be the only fields that were identified as being mandatory by almost all the respondents. The domain field was the next highest mandatory field with only 38%. Therefore, we can conclude that in most cases a term record containing only source and target terms would be

acceptable for this survey's respondents.

In 2005, LISA conducted a survey on terminology management where respondents ranked the information they collected for a term as shown in Table 10.

LISA 2005	%
Categorical	81
Contextual	78
Semantic	78
Administrative	65
Relational	50
Grammatical	45
Illustrations	18

Table 10 LISA 2005 Results on Recorded Information

The difference in categories and the fact that in our survey respondents were asked to indicate whether fields were required or mandatory create too wide a gap between the two sets of data to make a direct comparison of the results of both surveys possible.

The types of fields included in the records created by respondents to this survey coincide with the fields proposed by terminologists. For some sample bilingual records found in terminology works please refer to section 2.2.3.

The largest difference is in which fields are considered mandatory, i.e. must appear and be completed in all records. On Dubuc's proposed term record, all fields are mandatory (2002, p. 82). This position is very far from the results gathered in our survey, which indicate that the only fields that are close to being considered unanimously mandatory are source and target term. Other terminology approaches identified in the literature are more flexible. In the record structure established by the *Grand dictionnaire terminologique*, some fields are optional given that not all terms will be associated with a sub-domain, an officialization or sub-entries such as synonyms, quasi-synonyms, abbreviations, etc. (OQLF, 2002). Pavel and

Nolet clearly state that “[r]epetition of information in the textual supports is avoided whenever possible” (2001, p. 48). This leads the authors to suggest that terminologists are not required to provide both a definition and a context if the different supports do not offer additional information (Ibid.) These last two record structures consider many more fields to be optional, as seems to be the preference of our survey’s respondents. However, these term record structures require a definition and/or a context as well as sources, which are not considered mandatory by a majority of our survey’s respondents.

Finally, if we compare the results collected in our survey to the results of Durán Muñoz’s (2010, pp. 9-10) survey on translator’s needs with regards to lexicographical resources, we see that although the results differ in terms of the information that is considered mandatory as opposed to optional, the results of the two surveys conclude that grammatical and semantic information is not a priority while linguistic and pragmatic information is of greater interest.

The differences between what is considered essential as opposed to desirable may be owing to the fact that integrated termbases are self-created resources while lexicographical resources are reference works created by a third party (e.g. general or specialized dictionaries, on-line glossaries). For a self-created resource, we can guess that respondents are more lenient with what information must be included as they can easily decide what information can safely be omitted for their own future reuse of the resource, while for a reference resource they may prefer having that information at hand to be able to confirm that the information found is adequate to the context desired or to validate the reliability of the resource.

Essential Data	Desirable Data	Irrelevant Data
Clear and concrete definitions	A great variety of units (n., v., adj.)	Etymological information
Equivalentents	An explanation of each translation equivalent	Pronunciation
Derivatives and compounds	A greater variety of examples	Syllabification
Domain specification	Grammatical information	
Examples	Semantic information (semantic relations, frames)	
Phraseological information	Pictorial illustrations	
A definition in both languages (if bilingual) (45.11%)	A definition in both languages (if bilingual) (45.11%)	
Abbreviations and acronyms	Instructions for use	

Figure 13 Reproduction of Durán Muñoz’s table “What do you think a good terminological resource for translators should offer?”

5.5.4.2 Filling out a term record template

In this next section of the survey, respondents were presented with a list of types of units and were asked which of the units they would record as main entries for a term record. The only type of unit that more than half of the respondents accepted as a main entry was the full form of a term (92.8%). Far less popular alternatives were abbreviations (49.3%), acronyms (47.8%), phrasemes (36.2%), preferred spelling and base form (both at 23.2%), alternate spellings (17.4%), alternate forms (13%), symbols (10.1%), initialisms (8.7%), and URLs (2.9%). No one considered telephone numbers, fax numbers, email and physical addresses as information to be recorded as a main entry.

Synonyms can be recorded in several different ways within a termbase and the results reveal that users do record them in different ways. Results indicate that 37.1% of respondents record all synonyms of a concept as terms within a single record, 25.7% of respondents record synonyms in separate records without adding cross-references, 17.1% include synonyms in a supporting field of the term that is considered to be the main entry, and 15.7% record synonyms in separate records and indicate each term’s synonyms in a

supporting field.

When users are faced with synonymous units, often there is a preferred term. Of our respondents, 43.7% report indicating which one is the preferred form while 39.4% do not.

Units can have several morphological forms depending on the gender, number or case. Most of the respondents always record units in their base form (44%). The second most popular approach is to record the form that the user considers to be the most frequent (34.7%), and a distant alternative is to record the unit in the form in which it appears in the text being translated (17.3%).

Moreover, units may also occur as part of a collocation or combination of words. A majority of the respondents tend to record the unit with its collocate or combination (65.3%) and to do so in the main entry field in the record. In most cases, respondents record these combinations strictly, including only the determiners or adjectives with which the unit appears in the text (72.3%); only in some cases do respondents record these units with the determiners and adjectives they consider most likely to occur with the unit (27.7%). When recording this type of unit, respondents do so mostly in the base form (61.7%) or else in the form in which they appear in the text (42.6%). Respondents will rarely record all the gender, number or case forms possible (each option received a 2.1% response). In the event of recording multiple forms of a combination, respondents tend to record them in separate records (61.7%).

When respondents search for potential collocations or forms of an expression, they mostly turn to their Internet search engine (60.8%). Other tools they may use to further their research include paper dictionaries and thesauri (51%), their TEnT search function (47.1%), a concordancer (27.5%), the search function in a word processor (23.5%), a web concordancer (15.7%), an electronic database of words conceptual-semantic and lexical

relations (e.g. WordNet, Visuwords) (13.7%), or the search function of a document management system (7.8%).

5.5.4.2.1. Results in context

Established terminology methods prescribe a series of principles regarding how terminological information should be recorded. For example, according to the literature, units added to the main entry field must be terms and can be accompanied by a synonymous form if one exists (Dubuc, 2002, p. 82; Pavel and Nolet, 2001, p. 50; L'Homme, 2004, p. 39). Synonymous forms may include abbreviations, spelling or syntactic-variants, quasi- and pseudo-synonyms (Pavel and Nolet, 2001, p. 50). Such units are recorded in their base form, i.e. in singular, in masculine if the language inflects, in lower case unless the unit commonly appears capitalized, etc. (Dubuc, 2002, p. 82; Translation Bureau, 2008c).

Research on terminological practice among users of TMSs integrated within TEnTs reveals the following observation: given that the purpose of this type of tool is to identify terms from the source text and (semi-)automatically replace them, users may be inclined to break free from strict terminological principles and record term units either in their most frequent form or in several of their forms all in separate term records in order to facilitate automatic insertion into the target text (Kenny, 1999, p. 74). Bowker (2011, p. 223) complements this observation by adding that synonymous forms may also be recorded in separate records. This reflects a switch from the strict onomasiological approach traditionally favoured by terminology theory to a more semasiological one.

The results from our survey indicate that the most widely accepted form for the main entry is the full form of a term, as established terminology theory dictates. However,

although terminology theory indicates that abbreviations and alternate spellings should be entered in the main entry field given that they are synonymic forms, only 23% of respondents would include abbreviations in the main entry field and, 13% of respondents would include alternate spellings.

As far as synonymous forms are concerned, practice is divided between the approach grounded in terminology theory, with 37.1% of respondents entering all synonymous forms in the same record, and the semasiological approach pointed out by Bowker (Ibid.), with 41.4% of respondents creating separate records for each synonym.

Finally, among our survey's respondents, recording alternate forms of a unit is a rare practice, with only 4% of respondents doing so. However, as indicated by Kenny (Ibid.), although translators continue, by and large, to follow terminology theory principles and record the base form when recording such units (44%), they are also starting to adapt their practices in order to optimize their tools by recording units in the forms they consider to be most frequent (34.7%) or in the form that appears in the text (17.3%).

5.6 Discussion

After presenting our survey results and putting them in context with existing literature on the topic, this section will present the conclusions on the first two parts of the survey. The conclusions on the last two parts of the survey, which focus more directly on current practices vis-à-vis the preliminary hypotheses, will be presented in the next chapter, which will concentrate on the evaluation of the preliminary hypotheses and identification of hypotheses that require further testing.

5.6.1 Respondents' profile: professional

As presented in section 5.5.1, our survey results reveal that the most common source language was English. This can be attributed in part to the predominant role of this language worldwide, and, in part, to the requirement for all survey respondents to have a working knowledge of English. Target language distribution was more varied yet still dominated by English. This can be explained by the dominant status of this language across the world and its well-known role as *lingua franca* in most international business, research and leisure exchanges.

Most responses to our survey were received from translators, which is consistent with other survey results and may be related to the fact that translators have been targeted as the main users of TEnTs. However, as illustrated in section 5.5.1, reports (e.g. Champagne 2004a, p. 19; CTISC 1999, pp. 4, 20) concluded that there is a far greater number of translators than terminologists, project managers or any other occupations within the translation and localization industries. Therefore, the results of our survey seem to be in line with what can be deduced to be the actual market distribution of project managers, revisers, terminologists and translators. However, as noted above, it is very hard to determine the exact number of translators, terminologists, project managers, revisers and other job occupations within the industry because the different statistics agencies frequently combine these occupations under the language profession group. Therefore, calculations can only be obtained by approximation.

As for the respondents' breakdown by type of employer, as mentioned above, it should not be inferred from these results that most users of TEnTs in the market are freelancers (Champagne 2004a, p. 19; CTISC 1999, pp. 4, 20). This result may be also influenced by the mode used to distribute the survey: TEnT online user forums. The

independent nature of freelance work forces these users to be their own IT technicians, and they often tend to turn to TEnT forums to find the help and advice that translators working as part of in-house teams may find through their colleagues or the company's IT department.

With regard to the fields of specialization, it may at first glance seem surprising to see marketing listed as the third most common field of specialization, given that we usually associate those types of texts with creative and non-repetitive writing. However, branding is a central aspect of marketing and it relies on always preserving a company's terminology, slogans and style. One of the principal strengths of a TEnT is maintaining and encouraging consistency, and this may explain why these texts are also being translated with the help of such tools.

Finally, the results obtained about the age range of respondents shed a different light on the average TEnT user than we might have expected. Whenever computer programs and information technology are discussed, we tend to associate them with the young and "technologically hip". When that discussion focuses on a software tool that did not become widely available until the mid-to-late 1990s, that assumption is likely to become even stronger. However, according to this survey that common impression may mislead us when it comes to identifying typical users of TEnTs, since almost 73% of the respondents were over the age of 35.

In a closer look at the results per age bracket, the group of 18 to 24 year olds accounted for only 2.2% of the responses. The likely reasons for this are the following. Firstly, according to Statistics Canada (2008), the average age of graduation from a Bachelor's degree program is 22 years old. Some universities, such as Brigham Young University (Utah, USA), report that students graduate from a Bachelor's program outside of

this age bracket, at 25.4 years old on average. This reduces the pool of potential respondents within this age range. Secondly, although the invitation for this survey was posted in forums and distribution lists open to translators of all ages, the individual invitations were sent to personal acquaintances all over 26 years old. This may have restricted the number of invitations that reached younger translators as most likely the acquaintances of our acquaintances belonged to the same age bracket.

The limitations on the sample size and approach do not allow us to conclude that the vast majority of TEnT users are over 35 years old or that 30% of users are over the age of 50 (See Figure 14). However, these results clearly indicate that we should not assume that all TEnT users belong to the generations that were born post-WWW. TEnTs have been adopted not only by translators who arrived on the market after the popularization of this tool but also by experienced translators who have learned and acquired this new technology as it has been introduced in the industry. We can only guess the reasons that encouraged translators of all ranges of experience to adopt the use of a TEnT as the survey did not cover this question. On the one hand, respondents who adopted the use of a TEnT of their own choice most likely did so to remain competitive (e.g. to be able to be more productive, to offer discounted rates on repetitions or to provide clients with the assurance that translation technology will be used to ensure consistency and quality checks). On the other hand, respondents may have been requested to adopt the use of a TEnT either by their employer, if working in-house, at a client's request, or to be able to bid on specific contracts, if working as a freelancer.

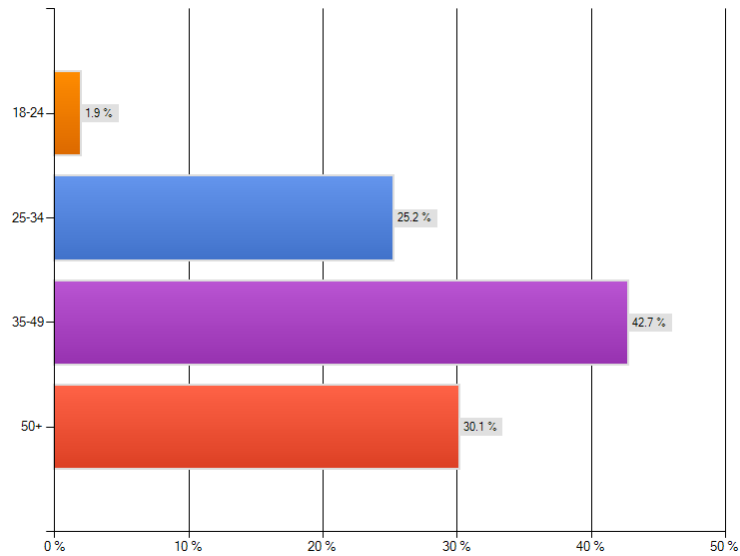


Figure 14 Respondents' Age Distribution

5.6.2 Respondents' profile: technological

When it comes to respondents' technological background, the predominance of SDL Trados does not come as a surprise. The German company Trados, founded in 1984, was one of the earliest TEnTs in the market with the release of their first Translator's Workbench dating to 1992 (SDL, 2009). Being one of the first in the market, Trados had a virgin market to explore. The sophisticated nature of this type of tool also ensured Trados a clientele that remained loyal over time due to the high cost involved in transferring resources between TEnTs, learning to master a new tool and assuming the diminished returns during the transfer and implantation period.⁵⁴ In addition, in 1997, Microsoft, who was already using Trados for its own localization needs, acquired part of the company, which gave the German TEnT an important boost thanks to the advantage of having access to Microsoft's marketing and distribution resources (García, 2004). In 2004, LISA's *TM Survey* established

⁵⁴ This type of situation is sometimes referred to as "vendor lock-in".

that Trados dominated 71% of the localization market (p. 12). This predominance was confirmed in 2006 by Lagoudaki's survey, reporting a 76% market share for Trados (p. 18). In 2005 SDL, ranking second in market share according to the results in both surveys mentioned above, acquired Trados to become the current SDL Trados.

An interesting observation when it comes to market presence is that the TEnTs WordFast and Omega-T, solutions traditionally offering fewer but more specialized features at a more affordable price tag or even free of charge, appear to have a relatively strong presence among users. An advantage of solutions such as WordFast or Omega-T is that they are multiplatform (both are compatible with Windows, Mac and Linux operating systems) and hence can appeal to those translators who would prefer alternative platforms to Microsoft. Most other main tools, including SDL Trados, Déjà Vu and MultiTrans, are compatible only with Windows as an operating system⁵⁵. Moreover, the fact that these tools are very economically priced (WordFast costs €300 and Omega-T is available as an open-source tool) can explain their increasing popularity among translators. As Lagoudaki (2006, p. 24) points out, most freelance translators tend to have smaller budgets and therefore find their choice of tool limited by the tool's price tag.

Previous surveys had already established that translators use multiple TEnTs, and this is confirmed by our survey. Given the purchase price and learning investment required by TEnTs, it is unlikely that translators own more than one for reasons of mere curiosity. A reason that might motivate translators or companies to invest in these tools may be that they need a complementary tool to carry out certain tasks for which their main TEnT lacks functionality (file format incompatibility, terminology extraction, file export to a standard format, etc.). Alternatively, clients may request projects to be carried out with specific

⁵⁵ These tools can all be run on virtual PCs from a Mac operating system, but vendors do not provide technical support for this type of installation.

TEnTs in order to be able to easily append the resulting translated texts to their existing TEnT databases and so benefit from the editing function of a specific TEnT that will spare them having to reformat the translation into the desired file format (e.g. web pages, files created with complex desk top publishing tools). Depending on the volume or duration of the translation contract, it may be worth the initial monetary and time investment.

5.6.3 Respondents' profile: training

Results seem to indicate that courses delivered by the TEnT provider have become more common in recent years. The percentage of users who followed courses on TEnTs at their academic institutions is fairly low at 27.5%. However, this can likely be explained by the fact that close to 75% of respondents are 35 years of age or older and therefore probably completed their studies at a time when TEnTs had not yet, or had only recently, been commercialized and therefore may not have been included in teaching curricula.

With regard to the content of the training, some results were more predictable than others. On the one hand, it is surprising to see that the TMS was included in the course content in only 37.3% of trainings. This may indicate that the TEnT is perceived mainly as a translation tool and that there is a lack of understanding about the role of an integrated TMS. Alternatively, it could also be that terminology was purposefully not included in the course content when terminology is managed by a group of terminologists or by another department of the company (e.g. marketing). In an academic setting, this may be due to the fact that the TMS is sometimes covered primarily in a separate terminology course and not as part of a translation technology course (Bowker and Marshman, 2009, p. 66). Finally, some tools offer the TMS as a component that must be purchased separately (e.g.

MultiTerm in SDL Trados). In such cases, if a client has not purchased that module, it would be understandable for it not to be covered during training. Regardless of the reasons, there is still work to be done to highlight the relevance of terminology management within the translation process and the benefits to be gained by learning how to use a TMS integrated within a TEnT in order to optimize the overall performance of the TEnT.

On the other hand, the fact that 53.3% of the training sessions that did cover the TMS integrated with the TEnT did not address what types of units to record and how to record them is somewhat understandable. This could be owing to a lack of awareness about the relevance of terminology in translation or within a TEnT, if the trainer has a technology or business background rather than a translation or terminology background, for example. In addition, as mentioned before, the lack of literature specifically focused on terminology management within TEnTs does not help educators who are new to the industry to develop this awareness. Even for educators who are well aware of the role of terminology in translation and within TEnTs, the paucity of available literature represents an obstacle as it forces them to develop their own content, either adapting the available terminology theory or relying on their own experience. Hopefully, the research resulting from this current project will help educators to fill those gaps.

Regardless of the lack of literature on the topic, it remains significant that 70% of the courses provided in academic settings did not cover what units to record within the TMS integrated with a TEnT. The high percentage may indicate other underlying reasons. Firstly, it could be that in some university programs, such as the one at the University of Ottawa's School of Translation and Interpretation, TEnTs and TMSs may typically be covered in separate courses such as "Introduction to Terminology and Terminotics" or "Translation Technologies" rather than being integrated into general and specialized translation courses

(Bowker and Marshman, 2009, p. 67). These tool-centered courses cover different types of tools (e.g. TEnTs, TMSs, term extractors, concordancers) and their functions but may not be the best context in which to put these tools into practice in real scenarios, where students may address not only what these tools can do but how they can best be used. Another reason that may have led to courses not covering what units should be recorded by translators may be the fact that these courses separate TEnT tools into terminology management tools and translation tools (Bowker, 2011, p. 24; Bowker and Marshman, 2009, p. 67). Thus, on occasion the link between the two TEnT components and the contribution that both make to the translation environment may be less evident, if they are presented as separate tools and the emphasis is not placed on the tool-suite aspect of the software. In addition, another possible logical explanation for the lack of focus on what units may most usefully be recorded within a TEnT may be the fact that these tools are sometimes introduced in terminology courses aimed at training future terminologists rather than translators. Such courses are more likely to focus on the question of what constitutes a terminological unit rather than on what units can be valuable to record when creating an integrated termbase for the purpose of translation. Therefore, in those courses, term selection criteria may be discussed more from a thematic research perspective. In such courses, the question of what units can be most relevant within the context of a TEnT may be overlooked or seem tangential. Finally, advising on best practices requires an expert knowledge of TEnTs and TMSs. According to the *eColore Translator Training Survey*, only 25% of respondents felt completely confident teaching TMSs to others (2006, p. 15) and 32.7% felt that way about teaching CAT tools (2006, p. 14).

Hopefully, the present research will help to promote the need to increase the awareness not only of the importance of understanding that TEnTs are tool suites whose

components are interdependent, but also that terminology management for translators and within TEnTs is a practice that has elements in common with thematic terminology research but which also has specific needs that will only be understood when these tools are used during translation practice.

5.6.4 Respondents' profile: perception

The general awareness and appreciation of the need for terminology management among our survey participants seems to be a little at odds with the fact that only 30% of respondents consider that they have mastered their TMS in full. This may be due to several reasons. For instance, it could be that basic terminology management features (e.g. creating termbases and term records and looking up information) meet the day-to-day needs of the survey respondents. However, advanced features are typically the ones that allow the user to carry out maintenance tasks on the databases as well as to import, into the TEnT, glossaries and terminology files originally in different formats.

Not being able to use these advanced features may come at a cost in the future. Firstly, carrying out maintenance is a key process to ensure that databases do not contain inaccurate, out-of-date or scattered information that would hinder the terminology search instead of facilitating it. Secondly, being able to import terminology files provided by clients or other reliable sources and then using the TEnT technology to ensure consistency and validate adequate terminology use forces the user to pay attention to terminology as part of the translation process and facilitates the task of the translator, who no longer needs to consult multiple glossaries to ensure that the right term is being used. The negative side effects of not mastering the advanced features have a greater impact on freelance users

because they cannot rely on terminologists, project managers or colleagues with more TEnT experience to assist them with these tasks.

Another reason why relatively few of the respondents feel that they have mastered the TMS integrated with their TEnT may be related to the type of training received. We must bear in mind that nearly 50% of the respondents were self-taught and that only 23.5% of the self-taught respondents considered themselves to be expert users of their respective TMSs, compared to 36.7% of respondents who received formal training. Learning software on one's own can prove more difficult, particularly with regard to advanced features, because their functioning and usefulness may be harder to grasp in the absence of guidance. Moreover, even when respondents had received formal training, in 37.3% of cases the TMS was not part of the course content. On the one hand, this indicates a need for courses focusing on intermediate users who are ready to tackle the more complex features of their respective TEnTs, including those of the terminology management component. On the other hand, it also highlights a need for guidelines on the relevance of the terminology management component within the TEnT and how to best build and use it, so that trainers and self-taught users can better understand this part of the tool and include it in their courses and everyday practice. As mentioned earlier, this research aims to fill this gap.

5.6.5 Terminology management planning: usage

The low number of responses to the questions in this section can provide only pointers to possible reasons that might discourage users from recording terminology. This section inquired whether additional features, training or documentation would encourage respondents to start recording terminology. These results suggest that from the users' and

educators' perspective there is still work to be done on teaching the relevance of the role of terminology management for translation and on learning how best to use the TMS integrated within a TEnT. TEnT developers may do well to investigate the user-friendliness and intuitiveness of their TMSs as well as with the degree to which they are integrated within the TEnT.

A recurrent claim in the literature is that users keep their terminology in spreadsheets, databases or word processor documents rather than in a TMS to facilitate data exchange. While exchangeability was certainly an obstacle a few years ago when each TEnT used its own proprietary format, the arrival of the XML-based standards developed by LISA — first TMX (Translation Memory eXchange) and now TBX (Term Base eXchange) — has gone a long way towards standardizing data transfer.

TMX was initially created to facilitate the exchange of TM databases across different TEnTs, and over time TEnT providers have also adopted this format to share termbase information. While using TMX files to exchange data eliminated the barrier of the proprietary format, the obstacle of the term record template remained. The user providing the termbase had to share the original term record structure of the database and the recipient was limited by this structure when importing the shared data. This required that the target TEnT be able to recreate the source term record structure for a successful transfer. If any characteristics of source term records (type of fields, number of fields, field names, etc.) could not be reproduced in the target TEnT, some data had to be excluded at best, or the transfer may not be possible in the worst of cases.

TBX is not only a standard encoding in XML, as TMX was, but it also establishes a default term record structure⁵⁶ and its goal is to eliminate the problems listed above. Thanks to the use of a common record structure or, more precisely, thanks to TEnTs being able to export to their termbases to TBX and import from this format, sharing terminological information across different TEnTs should be considerably more straightforward.

As it stands today, the widely distributed TEnTs in the market support TMX and although not all of them are TBX-compatible, this standard is rapidly imposing itself as the essential requirement it was designed to be⁵⁷. A reason why this standard is becoming more popular in the industry is because it has now been published by the International Organization for Standardization as ISO30042:2008 “Systems to manage terminology, knowledge and content — TermBase eXchange (TBX)” (ISO, 2008).

As this new standard becomes more widely available, industry associations such as the Localisation Research Centre (LRC), educators and TEnT providers will need to continue to promote the existence of this standard and its benefits.

5.6.6 Terminology management planning: Guidelines

The literature review contained a detailed discussion of the risks associated with not managing terminology. One of the paramount risks is introducing terminological inconsistency into documentation, which can increase the time and effort required for correction (Translation Bureau, 2005, p. 31; Dunne, 2007, p. 37; Lommel, 2005, p. 3), diminish productivity (Translation Bureau, 2005, p. 8; Lommel, 2005, p. 3) and weaken a

⁵⁶ LISA has established different types of TBX standards of varying complexity to cater to the different needs users may have.

⁵⁷ SDLTrados 2009 (SDLTrados, 2009), Star Transit ^{NXT} (Star, 2009) and MultiTrans (MultiCorpora, 2010) are currently compatible with TBX, and Déjà Vu provides a TBX template on which to base term records (Atril, 2003).

company's brand image (Translation Bureau, 2004, p. 26; Dunne, 2007, p. 37; Fidura, 2007, p. 41). It was therefore not surprising that 52.6% of respondents to our survey considered the role of the TMS within their TEnT to be either important or very important.

What is surprising, however, is that 53.2% of respondents do no planning at all with regard to the information that is recorded in the TMS. An unmanaged termbase can be as dangerous as – or even more dangerous than – not managing terminology because it may create a false sense of reliability when there is no real quality control being applied to the termbase contents. Therefore, a translator may erroneously trust an equivalent proposed by the system, or the system may distract and hinder the translator by proposing multiple matches of various degrees of relevance and reliability, thus forcing him or her to do extra research or, at least, to sift through and consider the various options proposed.

Freelancers are the group that puts the least amount of planning into how to design and build their termbases with 69.4% not having any guidelines in this respect. Because in their case they are typically the sole users of the termbase, the practice of not having an established process for how to organize it or feed it likely derives from the fact that they would make decisions on a case-by-case basis, and having built the database one record a time, they would be extremely familiar with its content. This approach can work very well when translators work with a small termbase, a highly delimited field of expertise, and a restricted number of clients, or when one has a flawless memory. However, when the termbase grows in size, or when translators work in multiple domains or with a large number of clients, the chances of remembering the nuances of each record or even its origin may become a challenge that risks becoming costly. It is for this reason that any termbase, regardless of the initial size and number of users, should be created and built in an organized and systematic way.

These pitfalls increase exponentially with the number of users sharing a termbase. Since up to 33% of multi-member in-house teams use no guidelines to help them create and build their termbase, the results indicate that there is clearly still a long way to go to improve the awareness of the importance of managing terminology within the translation industry.

Another question that comes up when faced with these results is whether users are not creating and following guidelines because they are oblivious of the risks they incur or because they do not know how to tackle such a task. Our research aims to investigate this issue on both fronts by raising awareness within the translation community about the importance of managing terminology, and then exploring ways to optimize this practice.

A final point is whether respondents considered specialized courses organized by academic institutions or by a professional association or industry organization to be important when creating their terminology management guidelines. In our survey, the division of opinions regarding the usefulness of these courses may be related to the fact that only 62.7% of training sessions covered terminology management and only 46.7% of those went so far as to discuss the type of content to record. Therefore, courses that covered terminology management and went into some depth about how to organize and feed a database may have generated a very positive impression, while courses that did not address this subject may easily have disappointed users in this regard.

5.6.7 Terminology management planning: Database organization

The results in this section of our survey revealed interesting facts regarding how users organize their databases and what use they make of them.

Firstly, the fact that most users keep multiple termbases seems to conflict with the

advantages promised by the introduction of TEnTs that allow translators to centralize their linguistic assets. There is a series of disadvantages inherent in keeping multiple termbases. For instance, termbases take time and effort to develop. Keeping multiple separate termbases risks placing terminological data in information silos, thus diminishing the chances of reuse. In addition, keeping several termbases multiplies the maintenance obligations, i.e. the time required to keep termbases clean, up-to-date, organized and complete.

There may be different reasons motivating translators to keep multiple termbases. Translators are very aware of the risks that mixing client or subject field terminology can entail. Such “train wreck” (Topping, 2000, p. 60) risks include poor-quality texts, client dissatisfaction, and message inaccuracy coupled with a variety of civil liabilities if the inaccuracy were to endanger a human life or a business’s profit margin or image. Keeping different sets of terminology separate is indeed a useful strategy for avoiding such undesirable consequences.

However, TMSs offer database organization tools such as classification by client, project, domain, and user, and these can be paired with filters and exports that allow users to properly separate units within a single termbase with the same efficiency as keeping separate termbases. The results reveal that most translators choose not to use these features in their TMSs. This may be related to the fact that only 30% of users consider that they have mastered the advanced features of their TMS, and that this approach requires translators to be disciplined in properly classifying terms within the termbase.

Secondly, the correlation between work settings and the personal or shared nature of respondents’ termbases revealed—unsurprisingly—that freelancers tend to use more personal termbases while in-house teams use shared ones. Nevertheless, this correlation did produce some thought-provoking results.

Results revealed that 18.4% of freelancers built some shared and some personal termbases and as many as 15.8% built exclusively shared termbases. The *Results in context* discussion of the *Database design* portion of section 5.5.2 described how although unexpected, these results may arise from the new arena for interaction that the Internet provides freelancers.

Regarding multi-member in-house teams, our survey revealed that one out of four respondents belonging to such a team maintains both shared and personal termbases. In this context, personal termbases may be used to record units that would not satisfy the team's guidelines, that may not be of interest to the majority of the group, that indicate a user's subjective preference or that contain personal notes or reminders.

Finally, small multi-member in-house teams of 2-9 members reported that in 11% of cases exclusively personal termbases were used. At first glance, it seems odd that a team of translators who work for the same employer do not share their resources. However, server technology that allows users to simultaneously share termbases comes with an expensive price tag and small teams are more likely to have a limited budget. It is possible that the use of personal termbases in this context derives more from the resources available than from a resistance to pooling efforts. The other possibility is that in a small translation team there may be less overlap in the translators' areas of specialization. If, instead of having 2 to 3 translators working interchangeably on similar types of documents, the team were to include a legal translator, a technical translator and a generalist translator, then having separate databases might be a desired approach and not a decision resulting from budget limitations.

6 Evaluation of preliminary hypotheses

This section will evaluate each of the preliminary hypotheses against the results obtained in our survey and the existing literature in order to determine whether the data can be considered to support the hypotheses strongly enough such that no further investigation is warranted.

6.1 Sub-hypothesis A

Contrary to what current terminology and terminography literature recommends, translators will use fewer term record fields in a TEnT-integrated termbase.

As discussed in section 3.1, our survey addressed this sub-hypothesis by asking participants which fields they include in their term record structure. More precisely, in answering this question participants had to classify a series of fields as mandatory, optional or not included, according to whether or not these fields appeared in their own records and whether or not they were mandatory. While there was no general consensus on the overall term record structure, two fields produced a virtually unanimous consensus: participants consider source and target terms mandatory fields (respectively receiving 98.7% and 96.1% selections). The next highest fields selected as mandatory were author (38.6%), domain (38%) and administrative information (e.g. client, project date) (35.6%). With regard to optional fields, participants were largely in agreement about several fields with more than 65% response rates: comments (81.5%), synonyms (78.5%), definition (69.9%), short forms (68.3%) and context (67.1%).

As previously discussed, the literature already pointed towards the fact that translators required fewer fields than terminologists. Examples of these claims are O'Brien's

(1998, p. 118) findings that showed that localization industry glossaries require only source and target terms or the fact that LISA developed the TBX-Basic standard, which constitutes a simplified version of TBX designed specifically for translation purposes. To this may be added the recent discussions about a newer TBX by-product called TBX-Glossary (Wright et al., 2010): it is another simplified TBX-compliant record structure, this time geared to facilitate the exchange of terminological information across tools compliant with a variety of existing standards.

Having analyzed the results of the current survey, which proves to be in line with the existing literature on the topic, the hypothesis that translators will use fewer term fields in TEnT-integrated termbases than terminologists do in term banks can be considered confirmed. For this reason, no further testing will be carried out.

The question that follows from this finding is which fields translators should retain. The description of the results obtained for section 5.5.4.1 illustrates in detail the fields that participants in this survey consider optional but that are nonetheless widely accepted. These results coincide fairly closely with the term record structure promoted by terminographic principles and by TBX-Basic.

Terminographic principles and the TBX-Basic record model are very valuable starting points for translators. Investigating which specific fields should generally be kept as optional and which should be excluded can be a futile exercise as every translator or translation service will have very specific needs that can justify the inclusion of a field or make it redundant. For example, for most users a product code may be of little use when they translate technical or marketing documentation. However, for someone translating a catalogue or localizing an inventory management application, product codes may be essential.

Researching which are the most valuable or necessary fields for specific translation contexts or for a specific use of a termbase may be a very productive research path. However, such an endeavour falls beyond the scope of this project.

6.2 Sub-hypothesis B

Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard.

This sub-hypothesis was first introduced in section 3.2 and TBX is described in the introduction to chapter 3. As indicated in section 3.2, although the survey did not address this sub-hypothesis directly, the section that investigates which fields are considered mandatory, optional or excluded sheds light on whether participants' current practices coincide with TBX-Basic. Based on our survey results, we can see that participants' practices follow certain TBX-Basic rules: a) source and target terms are mandatory fields and b) most of the fields considered optional are present in the TBX structure.

At the same time, our survey results contradicted other TBX-Basic rules. For example, in TBX-Basic either the part of speech (POS) or a context or definition must be present in a term record. The POS is required if the glossary is meant to be used as a resource for another application such as a machine translation engine. However, if the principal users are humans, either a context or a definition is required. According to our survey results, participants are divided between considering the POS field as optional or excluded, while the context and definition fields are generally accepted as only optional. These results are largely in line with O'Brien's findings which point out that in the

localization industry translators created glossaries including only the source and target terms (1998, p. 118); however, our respondents tend to include optional fields in their record structure.

Although current practices do not fully line up with the term record rules established in TBX-Basic and this standard would require translators to record additional information and work with a more complex term record template, translators may decide to adopt TBX-Basic because it provides a ready-made term record template that abides by industry standards and thus facilitates exchangeability.

Further testing will be required to be able to validate or reject this hypothesis.

6.3 Sub-hypothesis C

Contrary to what current terminology and terminography literature recommends, translators will classify records in personal TEnT-integrated termbases first by client or project and only secondly by domain.

This sub-hypothesis was first presented in section 3.3. The survey addressed this sub-hypothesis by first asking participants whether they work with a single termbase or multiple termbases and in each case whether records were classified by subject, client, project, language combination or any other criteria. The question allowed more than one answer.

In the case of participants using a single termbase, the question received 14 answers of which 64% classified records by subject, 50% by client and 57% by project. Meanwhile, 62 respondents used multiple termbases, of which 71% classified these by subject, 51% by client and 23% by project.

Because subject (domain) is the most common classification criterion for users of either a single termbase or multiple termbases, these results would seem to favour the rejection of the sub-hypothesis that users of integrated termbases would classify their records first by client or project and then by domain.

A closer look at the responses from users of multiple termbases, which is the dominant trend, is warranted. If we break down the results by participants working as freelancers or as a part of a team, the results remain consistent: in both cases domain remains the primary classification criterion at approximately 70%. Client is the second-ranked criterion, with stronger results coming from participants belonging to a team (62%) than from freelancers (42%). The project criterion ranks third with approximately 20% for both groups.

The relatively high results of each of the options indicate that participants use more than one classification criterion. When examining how these criteria are combined, we also note that while the records are classified most frequently by domain, only 32% of respondents who use multiple termbases classify solely by domain and 21% classify solely by either client or project and not by domain. Moreover, in 39% of the cases respondents classify their records by domain as well as by client or project.

However, this does not apply equally when results are broken out according to respondents who are freelancers versus those who work within a team. In the case of the 29 team members who use multiple termbases, 52% classify their records by domain and by client or project, while only 17% classify solely by domain and 17% also classify by client or project only. In the case of the 33 freelancers, the results are quite different. Among this group, 45% of participants classify their records only by domain and not by client or project, while only 27% classify them by subject and client or project.

In conclusion, the preliminary sub-hypothesis stating that translators will classify records in a TEnT-integrated termbase first by client or project and only secondly by domain is rejected. The results indicate that the sub-hypothesis should be reformulated to state that client and project are also common classification criteria in a TEnT-integrated termbase.

Further research into the reasons motivating users to select certain criteria over others or to work with a combination of criteria would definitely be interesting. However, such decisions will likely have a very strong link to each user's specific set of clients and domain of specialization. Users with multiple clients may choose to classify by client and domain in order to be able to reuse some records from different clients when they belong to the same domain, while users with fewer clients in very different domains may choose not to work solely with a client classification.

Another interesting area to explore in more depth would be what domain classification system users apply. Terminologists working within term banks use very detailed domain and sub-domain structures. While domain remains the most common classification field among integrated termbase users, the level of granularity used in this field may be very different than that applied in term banks.

Based on the results, the sub-hypothesis must be corrected to:

Contrary to what current terminology and terminography literature recommends, translators will classify records in personal TEnT-integrated termbases firstly by domain but also by client or project.

In the future, further research could be carried out to discern the motivations behind the classification fields selected with a view to matching classification fields and scenarios of use. However, for the purposes of this project, this sub-hypothesis was not subject to further testing.

6.4 Sub-hypothesis D

Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.

This sub-hypothesis was first presented in section 3.4. Our survey addressed this question by asking participants how their records were organized. The question received 70 answers. Participants were provided with a set of options from which they selected the one that applied.

In 54% of cases participants organized records around a concept. In 37% of cases they organized records around a concept and recorded synonyms as terms within the same record, following terminographic principles. In 17% of cases records were organized around a concept and synonyms were indicated in a supporting field. This practice has also been accepted within terminography; however, it was more popular in paper glossaries. In electronic glossaries, synonyms are generally entered as terms mainly to facilitate searches, which are most often done on the entry field. In 48% of cases, respondents recorded each synonymous form in a different record. When recording synonyms in separate records, 26% of participants reported that they did not indicate the synonymous relation in the record and in 16% of cases they indicated it in a supporting field.

Based on the results of our survey participants are divided nearly equally in favour of and against the practice proposed by the hypothesis. We can have a closer look at results based on the work setting of the respondents. The group of freelancers maintains the split with 48% of respondents recording each synonymous form separately and 45% of respondents organizing their records around a concept. When it comes to translators working in a team setting, the split is not so even: 62% of respondents organize records around a concept while only 35% create separate records for each synonymous form.

Such divided results do not allow us to validate or reject the preliminary sub-hypothesis. The fact that such a substantial proportion of participants currently break with terminographic principles to record each synonymous form separately does suggest that the proposed strategy may be of interest. Further tests will be required to establish whether translators find the organization of term records by equivalent pair useful when working with an integrated termbase.

The current results are not sufficient as translators' practice could be influenced by factors other than the need to create a termbase structure that will provide the most efficient results when translating interactively. For instance, those organizing their term records around a concept may be influenced by their education in terminology or their experience with printed glossaries and electronic term banks. In contrast, those who create a term record for each form may do so because of the limitations of a TEnT that may not allow more than one term per record, or simply because they create the records as they come across any term of interest without considering what they have already entered in their termbases.

Further tests on this sub-hypothesis will allow us to examine whether translators prefer the results generated by a TEnT that is integrated with a concept-oriented termbase or a form-oriented termbase.

6.5 Sub-hypothesis E

Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.

As mentioned in section 3.5, a term is a unit designating a specialized concept (Cabr e, 1998, p. 149; Dubuc, 2002, p. 33). In terminography terms are generally considered to take certain forms: “morphological and lexical structures: noun (simple, derived, compound), verb, adjective, noun phrase, verb phrase, or adjective phrase” (Pavel and Nolet, 2001, p. 19). This sub-hypothesis proposes that translators may find it useful to record other non-term units in an integrated termbase.

To investigate this sub-hypothesis, our survey asked participants to choose the types of units they recorded in their termbases from a list that contained a mix of the morphological and lexical structures above and other units such as URLs, emails, physical addresses, telephone and fax numbers, common paragraphs and common sentences.

The conventional morphological and lexical structures were by far the most commonly selected units on the list. However, it was interesting to see that out of the 72 participants who answered this question, 33% recorded common sentences, 16% recorded common paragraphs and a few also recorded URLs (8%) and emails (1%). No respondents indicated that they record telephone and fax numbers or physical addresses.

These results suggest that some users do record non-term units. However, based on this information alone, it is not possible to determine whether the participants who do not record these units make this decision because they do not find such units to be useful or because they had never previously considered the possibility of doing so. Therefore, this sub-hypothesis requires further research.

6.6 Sub-hypothesis F

Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will not be opposed to extracting terms/units and equivalents from translated texts.

This sub-hypothesis was first introduced in section 3.6. As presented during the discussion of our survey results, participants responding to this question had to answer whether they used their own TMs and if these were the only resources they checked, one of their top resources, one of the resources they considered, a last resort or if they never used them. The question received 76 responses and 86% answered that they do use TMs as a resource to obtain term equivalents. Not only do most respondents use TMs but 40% of respondents consider them one of their top resources.

Based on these results, the sub-hypothesis can be considered validated and no further testing will be carried out within the scope of this project. However, it may be interesting in the future to deepen the research on this topic in order to find out what criteria influence the acceptability of translated texts: personal translations, translations by other members of a team, translations from reputable sources and so on.

6.7 Sub-hypothesis G

Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or their most frequent form(s).

This sub-hypothesis was explained in detail in section 3.7. Our survey investigated this sub-hypothesis by asking participants whether they recorded terms solely in their base form, in the form in which they appeared in the texts they translate, in the form they consider the most frequent, or in all possible forms. The question garnered 75 answers and 44% of respondents indicated they always recorded terms in their base form in accordance with terminographic principles. The response was divided, however, as 39% of respondents indicated that they recorded either the most frequent form (34.7%) or all the forms (4%). Finally, 17.3% of participants indicated they recorded the term as it appeared in the text.

The results obtained for this answer show that a substantial portion of participants do not feel bound by terminographic principles when choosing how to record terms in their records. However, with such divided results further tests will be required to validate or reject the hypothesis.

6.8 Summary of sub-hypotheses to test further

We evaluated the preliminary sub-hypotheses against the results of our first *Use of Terminology Management Systems Integrated with Translation Environment Tools* survey and concluded that the following sub-hypotheses will require further testing:

- B. Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard.
- D. Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.
- E. Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.
- G. Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or in their most frequent form(s).

7 User-acceptance test

Once we had evaluated the preliminary sub-hypotheses and identified those that warranted further research, the next step of this project was to undertake further testing. To do so, we carried out a second survey, the *Integrated Termbases Optimization Survey*.

According to the results of our first survey, the practices described in the preliminary sub-hypotheses were not at all in use or were accepted by only a minority of the respondents, or the responses were much divided. In the second investigation, rather than asking users what they did, we decided to ask them whether they would consider the strategies described in the sub-hypotheses listed above to be worthwhile. The underlying idea behind this is that the sub-hypotheses can present an approach that the users may not have considered previously or may not be currently applying, but it may nonetheless be an approach that participants may find to be useful when they are presented with the examples of this approach.

7.1 Target audience and criteria for participation

As was the case with the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, in order to obtain the highest number of participants possible the fewest limitations possible were applied. The survey had no regional, professional or minimum experience requirements.

As explained in section 0, the only conditions each participant was required to fulfill were the following:

- a) consent to participate in the survey
- b) be 18 years old or older
- c) have good reading comprehension of English
- d) use a TEnT.

Given that we had no control over the sample when distributing the survey, a demographics section was included within the survey to obtain information such as: country of origin, profession, work setting, main source language, main target language and terminology training. This information allowed us to describe the respondents' profile and to do sub-group analysis of the data that was collected, as deemed necessary.

7.2 Survey design

The survey contained a total of 22 multiple-choice type questions and was divided into three sections: sample screening, hypothesis tests, and demographic profile. The complete survey questionnaire can be consulted in Appendix D. The sample screening section consisted of four mandatory questions, reflecting the selection criteria described above. The section addressing the sub-hypotheses tests contained 12 questions focusing on the four sub-hypotheses to be tested: a) recording synonyms (three questions), b) recording non-term units (four questions), c) recording term forms (two questions) and d) using TBX-Basic (three questions). Finally, the demographics section consisted of the six remaining questions.

Compared to the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, this survey was much more linear. It required skip logic only to exclude participants if they did not meet the requirements to participate in the survey (see section 0 for more details), and to guide users to follow-up questions based on their answers to two of the survey's questions.

The goal of the survey was to present participants with different ways of recording certain units in an integrated termbase, and to ask them to select their preferred option based on the results they would obtain. To achieve this, we considered that the best question format was multiple-choice, usually consisting of a description of a scenario and of the possible results, and the question asking the respondent to select which result he or she would prefer to work with. Follow-up questions attempting to determine the reasons behind a choice were also in the multiple-choice format, and in these cases a list of all potential options identified was provided along with a free-text field for any additional option that may have been inadvertently omitted.

Participants received clear instructions to answer the questions not thinking about their own TEnT with its strengths and weaknesses, but with a generic TEnT in mind. Moreover, they were told not to indicate their current practices in their answer, but rather to select the proposed options that they would honestly consider to be most useful.

The wording of each question and of any instructional text used in the survey was chosen carefully to avoid any bias. Whenever deemed necessary, a free text comment field was provided. However, no comment field was allowed for the questions in which participants had to choose one option over another. This omission was intentional in order to force participants to make a selection. Being aware that each translation project, text, term, TEnT, and work setting will have its peculiarities, we felt that if a middle-ground

option were to be given, participants would legitimately gravitate towards it. However, as the survey aimed at uncovering what their preference would be when working with a generic TEnT and facing each of the scenarios in a hypothetical situation, giving participants no middle ground was thought to be the best way to elicit the answer that they find to be the best, in general.

Given the length of the first survey and the impact that this length appeared to have on the response rate, a conscious effort was made to only include essential questions for our conclusions when designing the second survey. This was no easy task given that, for each question, a host of follow-up questions came to mind to investigate the whys, hows and what ifs. While the first survey needed to be exhaustive owing to its goal of establishing a description of current practices, this second survey aims to find out whether or not the proposed strategies are acceptable. Therefore, we limited the questions to address this point and added only a few minimal follow-up questions for what we identified as key issues. Future research that continues exploring the nuances of each strategy may be very interesting.

Once designed, the survey was submitted to the University of Ottawa's Research Ethics Board for approval. The *Integrated Termbases Optimization Survey* was granted approval on June 21st, 2011 (see Appendix A).

7.3 Survey distribution

The *Integrated Termbases Optimization Survey* is an online survey that was designed, like the previous survey, using the online survey provider SurveyMonkey. This tool was retained based on the positive experience we had when creating and administering the first survey of

this research project. SurveyMonkey proved to be a flexible online survey provider offering multiple question types, survey format templates and the necessary data analysis tools (cross-tabbing and filtering functionality). In addition, the fact that we had already learned how to create and analyze surveys using this tool made this second experience smoother.

The invitation to participate in the survey was distributed in different ways:

- a) An invitation was sent to all 43 participants of the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* who stated they would be interested in a follow-up survey and who provided their email addresses.
- b) An invitation was posted on the TEnT user forums relating to several TEnT tools: Déjà Vu⁵⁸, Memo-Q⁵⁹, MultiTrans⁶⁰, Omega-T⁶¹, SDLX⁶² and Trados⁶³.
- c) An invitation was also sent to a series of professional associations within the industry asking them to forward the invitation to their members: the American Translators Association (ATA), the *Associació Professional de Traductors i Intèrprets de Catalunya* (APTIC), the Association of Translators and Interpreters of Ontario (ATIO), the Canadian Association of Translation Studies (CATS), the *Ordre des traducteurs, terminologues et interprètes agréés du Québec* (OTTIAQ) and the *Sección Autónoma de Traductores de Libros de la Asociación Colegial de Escritores de España* (ACETT).

⁵⁸ <http://tech.groups.yahoo.com/group/dejavu-l/>

⁵⁹ <http://tech.groups.yahoo.com/group/memoQ/>

⁶⁰ <http://tech.groups.yahoo.com/group/multitrans/>

⁶¹ <http://tech.groups.yahoo.com/group/OmegaT/>

⁶² <http://tech.groups.yahoo.com/group/sdlx/>

⁶³ http://tech.groups.yahoo.com/group/TW_users/

- d) An invitation was sent to 57 academic, personal and professional contacts who either would qualify to answer the survey or might have acquaintances who do.
- e) Barbara Inge Karsh published an entry on her blog, BIK Terminology, inviting her readers to participate⁶⁴ and Jost Zetzsche published a follow-up article on the 197th edition of the email newsletter Tool Kit-Basic Edition (July 29, 2011).

All invitations included a letter of information with the project details as well as the contact details of this project's researchers in case the recipients required additional information.

Invitations encouraged recipients to forward the information to anybody they thought might be interested in the survey. Therefore, it is not possible to establish how many people received the invitation.

The following sections will present a summary of the results gathered, followed by a discussion of the relevance of the results for each of the sub-hypotheses tested and an analysis of the relevance of the data supporting or rejecting these sub-hypotheses.

7.4 Survey results

A total of 159 people began the survey and 122 completed it in full. Therefore, 37 participants started the survey but did not finish. The completion rate of the survey was 76.7%, a sizeable increase compared to the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* which had a 66.7% completion rate. It seems most likely

⁶⁴ <http://bikterminology.com/2011/07/08/survey-by-university-of-ottawa/>

that the increase is related to the fact that this survey is much shorter, with only 22 questions as compared to the 69 in the previous one⁶⁵.

Of the 122 participants who completed the survey, 109 met the selection criteria. Most of the disqualified participants (11) were excluded from the survey for not being users of a TEnT. The remaining two disqualified participants were excluded for not accepting the consent form or for not having good reading comprehension in English.

This results analysis will include only the answers of respondents who completed the survey and fulfilled all the selection criteria.

7.4.1 Respondents' profile

Given that there was no control over who would receive the survey invitation, a demographics section was included in the survey in order to be able to establish a description of the profile of the respondents. At a later stage in the analysis, this information also allowed us to isolate groups of respondents to carry out additional comparisons by work setting, profession or past education.

The 109 respondents who completed the survey and fulfilled the criteria originated from 29 different countries. Distributing the survey online and publicizing it in the user-group forums, professional associations and specialized blogs and newsletters had the desired effect of reaching translators across the globe. Predictably, as it is the base for this research, the country with most respondents is Canada (24), followed by the United States

⁶⁵ A few participants were lost along the survey and this may be due to the complexity of the questions asked. Although, all questions had the simple structure of a yes/no question, their content required careful consideration. This was noted in the survey posts and articles that Barbara Inge Karsh and Jost Zetzsche kindly published advertising its existence and calling for users to participate. In Karsh's post she mentioned that the survey required "brain power" and Zetzsche's article commented that questions were "designed to make you think hard".

(12) and Switzerland (11), which were the only countries with more than 10 participants. Table 11 illustrates all the countries represented in the survey and the number of participants per country. Readers should note that 2 respondents skipped this question.

Country	Participants	Country	Participants
Canada	24	Portugal	2
United States	12	Uruguay	2
Switzerland	11	Chile	1
Germany	7	China	1
Argentina	5	Ecuador	1
France	5	Finland	1
Spain	5	Greece	1
Austria	4	Iceland	1
Belgium	3	Indonesia	1
Italy	3	Japan	1
Poland	3	Norway	1
United Kingdom	3	Slovenia	1
Brazil	2	Turkey	1
Czech Republic	2	Ukraine	1
Denmark	2		

Table 11 Participants per Country

Most of the respondents were translators (68.8%), although a few identified themselves as terminologists (7.3%) or project managers (6.4%). This question allowed respondents to specify any other profession if they did not identify with any of the three above. The “other” category collected 17 answers. Some of these respondents had a split role including either all of the above responsibilities (2 respondents) or a combination of translation and terminology (3 respondents). Others were professors of translation or translation and terminology (4 respondents). An interesting category emerged in this “other” group: the one of translation technology specialist (6 respondents). Finally, among the respondents there was also a manager, a reviser, a translation coordinator and a localization analyst.

Regarding their work setting, 56% of the respondents worked as freelancers and 44% worked in-house. It must be noted that 8 respondents chose to skip this question.

The main source language was English (59%), followed by German (19%), French (9%), Italian (4%), Spanish (3%), Danish (2%), Dutch (2%), Finnish (1%), Japanese (1%) and Norwegian Bokmål (1%). Two respondents did not provide a main source language.

It also seems logical that there was a greater variety of target languages. French led with 28%, and was followed by English (24%), Spanish (12%), German (11%), Portuguese (5%), Italian (4%), Polish (4%), Catalan (3%), Czech (2%), Chinese (1%), Dutch (1%), Greek (1%), Icelandic (1%), Russian (1%), Slovenian (1%), Turkish (1%) and Ukrainian (1%). Two respondents skipped this question.

Finally, as the survey focused on terminology management practices, it was considered important to find out whether participants had received any formal training (such as university courses, certificates or workshops) in terminology in the past. It turned out that 71% of respondents had received formal training in terminology while 29% had not.

7.4.2 Recording synonyms

The first of the content questions focused on preliminary sub-hypothesis D, which dealt with synonym recording: *Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.*

To test whether such an assumption would prove true, and in a case of term synonymy, whether respondents would indeed prefer the results obtained when a term is recorded by equivalent pair instead of by concept, several questions were designed.

The questions were divided according to the translation method. Two questions were created using the scenario of interactive translation and one of pretranslation. As a brief reminder, interactive translation takes place when the TEnT examines each sentence of a new source text and, sentence by sentence, proposes the matches found in the TMs and termbases so that the translator can assess, adapt, and eventually insert the accepted translation for that sentence. Pretranslation takes place when the user of a TEnT allows the system to automatically replace any source text segment or term with its top-ranked equivalent as found in the TM or the termbase.

Firstly, respondents were asked to assume that they would be translating a text interactively. In this scenario they were faced with a text that contained a term presenting a case of synonymy: *aspirin*. The concept “aspirin” can also be designated in a more specialized register as *acetylsalicylic acid* and this synonymy repeats in the sample target languages used in this example: French (*aspirine / acide acétylsalicylique*), German (*Aspirin / Acetylsalicylsäure*) and Spanish (*aspirina / ácido acetilsalicílico*). If synonyms are recorded by concept, all synonyms for a term will be recorded in the same record, and results for a sentence containing the term *aspirin* would be displayed in a TEnT as shown in option A of Figure 15. If synonyms are recorded by equivalent pair, each pair or set of equivalents (*aspirin - aspirine - Aspirin - aspirina* vs. *acetylsalicylic acid - acide acétylsalicylique - Acetylsalicylsäure - ácido acetilsalicílico*) would be recorded in two separate records and the results obtained for a sentence containing the term *aspirin* would be displayed in a TEnT as shown in option B of Figure 15.

Option A				Option B			
EN	FR	SP	DE	EN	FR	SP	DE
aspirin	acide acétylsalicylique aspirine	ácido acetilsalicílico aspirina	Acetylsalicylsäure Aspirin	aspirin	aspirine	aspirina	Aspirin

Figure 15 TL Synonym Display (Full Form) during Interactive Translation

After evaluating the proposed scenario, 74% of respondents preferred the results obtained when recording synonyms by concept and only 26% preferred the results obtained when recording synonyms by equivalent pair. One respondent skipped this question.

Option A				Option B			
EN	FR	SP	DE	EN	FR	SP	DE
GDP	PIB	PIB	BIP	GDP	PIB	PIB	BIP
	produit intérieur brut	producto interior bruto	Bruttoinlandsprodukt				

Figure 16 TL Synonym Display (Acronym) during Interactive Translation

Secondly, continuing with the scenario of interactive translation, respondents were faced with the same question but for a sentence containing a different case of synonymy: an acronym. In this case it was the term *GDP*, which stands for *gross domestic product*. The sample target languages remained the same: French (*PIB/ produit intérieur brut*), German (*BIP / Bruttoinlandsprodukt*) and Spanish (*PIB / producto interior bruto*). The results of a concept-based record were presented as Option A (as shown in Figure 16) and those of an equivalent-pair-based record were presented as Option B (as shown in Figure 16).

Respondents' answers with regard to the recording of synonymy in the form of acronyms revealed even stronger support for the concept-oriented approach (89%) rather than that organized by equivalent pair (11%).

Finally, respondents were asked to make a similar choice but for a scenario where a text would be pretranslated. The term selected to illustrate this case was *global warming* with several synonyms available in French (*réchauffement climatique / réchauffement planétaire / réchauffement de la planète*), German (*Erdenwärmung / globale Erwärmung*) and Spanish (*calentamiento de la tierra / calentamiento global*). This case of synonymy does not present a direct relation of equivalent pairs. All target language equivalents can replace the source term with no

pragmatic variation (e.g. no differences in terms of regional usage or register). Therefore no option to record the synonyms by equivalent pair was offered.

In the context of interactive translation, this case of synonymy does not create any difficulty. Given that the source term would be the same, all forms would be presented to the translator during an interactive translation regardless of whether the synonyms are all entered in a single record or in separate records. The translator may then select the preferred option.

In pretranslation, this case is more interesting as it raises the question as to how the TEnT decides which target equivalent to insert in the text. If all target equivalents are entered on the same record as terms, different decisions will result depending on the tool. For instance, the user may be able to identify a main term to be selected, the tool may insert the first target equivalent that was provided or in some cases the TEnT may not replace the term in question and either leave it untouched or ask the translator to select the target term on the spot. If a separate record was created for each form, during pretranslation, the TEnT could not replace the term in question and would either leave it untouched or ask the translator to select the target term on the spot.

	Option A	Option B
FR	The main human activities that contribute to réchauffement climatique are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to réchauffement climatique / réchauffement planétaire / réchauffement de la planète are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.
SP	The main human activities that contribute to calentamiento de la tierra are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to calentamiento de la tierra / calentamiento global are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.
DE	The main human activities that contribute to Erdewärmung are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to Erdewärmung / globale Erwärmung are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.

Figure 17 Equivalent Insertion during Pretranslation in cases with Multiple TL Synonyms

Results for the selection of the term equivalent to insert during pretranslation in cases of multiple equivalents can vary significantly from one tool to the next depending on its internal logic. In this case, and as shown in Figure 17, the options presented to respondents were A) to record all equivalents on a single record with each synonym entered as a term within that record (and to assume that the TEnT would either insert the first equivalent or the synonym identified as the main target equivalent) or B) to record all equivalents on a single record but to enter all synonyms as one term (i.e., in the same field) in order to have them all inserted in the pretranslated text as alternate options.

In this particular case, 64% of respondents preferred the results produced by option B, where all synonyms are recorded as one single term and all forms are inserted in the pretranslated text. Meanwhile, only 36% of respondents opted for option A, which involved recording the synonyms as separate terms within the record and inserting only one form in the pretranslated text. One respondent skipped this question.

This scenario goes beyond sub-hypothesis D, which focused only on whether synonyms would be recorded by concept or by equivalent pair. However, during the preparation of the survey this case of synonymy was deemed of enough interest to warrant an additional question. It should be noted that option B implies the violation of another traditional terminology management principle: term autonomy. Term autonomy is the principle that dictates that terms should be recorded on their own without articles, prepositions or any devices to indicate other possible forms or equivalents (Wright and Budin, 2001, p. 583).

The discussion on the relevance of these results with regard to the sub-hypothesis that was tested can be found in section 7.5.2.

7.4.3 Recording non-term units

This next section of the survey was designed to test sub-hypothesis E: *Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.*

In the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, participants were asked whether they currently recorded non-term units. A number of the respondents indicated that they record common sentences (33%), common paragraphs (16%), URLs (8%) and emails (1%). However, other non-term units such as telephone or fax numbers and physical addresses were not recorded at all by participants.

This section of the survey presented participants with scenarios in which various non-term units (an email, a civic address, a URL and standard text) occurred in a text and required different equivalents (i.e. were not identical) in the target translation. See Figure 18 for the examples used.

Non-Term	EN	FR
Email	onlinedonations@redcross.ca	donsenligne@croixrouge.ca
Civic address	Citizenship and Immigration Canada Case Processing Centre – Sydney P. O. Box 7000 Sydney, Nova Scotia B1P 6V6	Citoyenneté et Immigration Canada Centre de traitement des demandes C.P. 7000 Sydney (Nouvelle-Écosse) B1P 6V6
URL	www.travel.gc.ca/offices	www.voyage.gc.ca/bureaux
Standard text	The decision to travel is the sole responsibility of the traveler. The traveler is also responsible for his or her own personal safety. The purpose of this Travel Report is to provide Canadians with up-to-date information to enable them to make well-informed decisions.	La décision de voyager revient à chaque voyageur. Il incombe également à chacun de veiller à sa sécurité personnelle. Les présents Conseils aux voyageurs ont pour but de fournir des renseignements à jour pour vous aider à prendre des décisions éclairées.

Figure 18 Examples of Non-Term Units

When faced with each of these concrete examples, respondents this time around were clearly open to the idea of recording non-term units: 61% approved creating a record for an email address, 69% approved creating a record for the civic address, 62% approved recording a URL and 70% approved recording frequently recurring standard text.

The discussion on the relevance of these results with regard to the sub-hypothesis that was tested can be found in section 7.5.3.

7.4.4 Recording multiple term forms

Sub-hypothesis G also required further testing: *Contrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or their most frequent form(s).*

In the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, participants were asked whether they recorded only the base form of each term, several forms or all forms of terms. Respondents to this initial survey were much divided on this matter: 44% recorded the base form, 34.7% recorded the most frequent form and 4% recorded all forms.

In our second survey (*Integrated Termbase Optimization Survey*), respondents were presented with a text containing various forms of the term *marinate* (marinate, marinating, marinated) and were asked whether they would only record the base form or multiple forms including all or a combination of the different variations. Of 109 respondents, 59% opted to record only the base form and 41% opted to record multiple forms.

It appears that in our second survey, the option to record only the base form was more strongly supported than in the previous survey.

Those who indicated that they would record multiple forms were then asked whether they would record all forms that appear in the text, the most frequent form in the text or all forms of the unit they could think of. Of the 45 respondents who opted to record multiple forms, 53% opted to record all units they could think of, 38% opted to record all forms in the text and 9% opted to record the most frequent form that appears in the text.

The discussion on the relevance of these results with regard to the sub-hypothesis that was tested can be found in section 7.5.4.

7.4.5 Using TBX-Basic as a term record template

This final section of the survey investigated sub-hypothesis B, which had not been addressed directly by the previous survey: *Contrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard.*

In this section, respondents were presented with a description of the TBX-Basic record structure. The description covered the fields available at each level, the type of each field, the pick-list options available, the optional and mandatory fields and any allowable modifications to field names and field options. The description was followed by two sample records that were compliant with TBX-Basic.

Respondents were then asked, if were they to start a new termbase and their TEnT had a built-in TBX-Basic-compliant term record template, whether they would use this record template or create their own. A majority of respondents (72%) indicated they would use the TBX-Basic compliant template while (28%) would opt not to use it. One respondent skipped this question.

Those who answered that they would use the TBX-Basic template were asked about the reason that most influenced their decision. The most popular reason was the assurance that TBX-Basic would facilitate the exchange of termbases among tools that are TBX-compatible (42%). The second most popular reason was simply that TBX-Basic met the respondents' terminological needs (25%). After that, 19% of respondents would opt to use the template because it would be easier than creating one from scratch and 14% would do so because it was based on an industry standard and compliant with an international standard.

Those who rejected the idea of using this built-in TBX-Basic record template were also asked for their reasons. Among the 31 respondents who indicated that they would not use a termbase based on the TBX-Basic template, 26% refused because they did not agree with a part of speech, a definition or a context being mandatory, 19% because it did not include a field they require, 16% because it lacked pick-list values they require, 10% because they were required to use another record template structure to maintain compatibility for exchanging data with certain software, clients or institutions and 3% because they do not agree with the term field being mandatory. Of the 31 respondents, 25% cited "other" reasons. The common denominator in these responses, as in the standard options provided above, was the lack of flexibility, ranging from the inability to create more complex records (e.g. with a higher number of fields or with different fields and/or pick-list options) to the possibility of creating glossaries in the form of a simple list of terms.

The discussion on the relevance of these results with regard to the sub-hypothesis that was tested can be found in section 7.5.5.

7.4.6 *Sub-group analysis*

When preparing the survey, we considered the possibility that users in different circumstances might react differently to some of the questions. The main factors considered likely to lead to variation were identified as the work setting (freelance vs. in-house translators), education (having received formal education in terminology or not) and profession (translator, terminologist, project manager, etc.).

When results were filtered by work setting, only minor changes in the trends appeared for most questions. Some answers did move a few percentage points up or down but there were no instances that came closer to a split result. The only noticeable difference between these two sub-groups can be seen when looking at the question of recording only the base form of a term or multiple forms. Freelance translators were very split around this issue: 45% opted for the base form and 55% opted for multiple forms. In-house translators seemed more inclined to record only the base form (72%) rather than multiple forms (28%).

Similarly, when respondents were analyzed according to those who had received formal education in terminology and those who had not, there were not many major trend changes identified. Only two questions caught our attention while comparing these two groups.

On the one hand, when it comes to recording the base form of a term or multiple forms of a term, there was a less equal split than had been observed in the general results. Roughly two thirds of respondents who had received formal education in terminology (68%) indicated they would record only the base form. The results are almost inverted for respondents who had not received formal education in terminology as 63% of them would record multiple forms.

On the other hand, an analysis of this subgroup reveals a difference when it comes to the recording of synonyms as separate terms within a record or all forms as the main term entry for pretranslation purposes. When looking at the responses for those who had received formal education in terminology, 71% would record all synonyms as the main term entry in order for the TEnT to insert all alternatives when pretranslating. Among those who did not receive formal education in terminology, there is less of a consensus: 55% of respondents would record each form as a separate term within a record and 45% would record all forms as the main term entry.

Finally, filtering results by profession did provide interesting findings across more questions. This was especially true when filtering the answers of the terminologists' subgroup; however, it must be noted that this was a very small sample of only eight respondents.

The initial expectation was that, for each question, terminologists would choose the option that adheres to the conventional principles of terminology management. Therefore, it was no surprise to learn that they unanimously chose the concept-based approach for recording synonyms when translating interactively or that all but one indicated that when faced with multiple forms of a term they would record only the base form.

What did catch our attention was that, when pretranslating, seven out of the eight terminologists opted to breach the term autonomy principle and to insert all target language equivalents in the translation. This was chosen over the option of preserving term autonomy but having to either select only one form to be inserted or having the tool pick a form for the user.

It was also surprising to discover that the sub-group of terminologists proved to be open to revisiting the notion of what might constitute a unit worth recording in a termbase:

five out of eight would approve recording an email address that changes in the target language, five out of eight would accept recording a civic address that changes in the target language and four out seven would record a URL that changes in the target language. However, the line was drawn in the case of standard text, where six out of eight respondents judged that it would not be useful to record in an integrated termbase a standard text segment that changes in the target language.

Finally, this small sub-group of terminologists seems receptive to the idea of basing an integrated termbase on the TBX-Basic structure when a template is already provided within the TEnT. Six out of eight responded that they would use this structure to design their termbases with the main reason for this being the guarantees for exchangeability offered by the standard. The remaining two terminologists rejected the use of the TBX-Basic template; one did so because it lacked a pick-list option that was deemed essential and the other because it is not suitable for complex environments such as the one this user faced (although the user would recommend it for other users with less elaborate terminological needs).

7.5 Discussion

Now that the data collected in the survey have been presented, this section will analyze the results and contrast them with the sub-hypotheses that the survey aimed to test. This section is divided into five sub-sections. We will first offer a discussion of the sample of respondents, followed by an analysis of each of the sub-hypotheses tested: recording synonyms, recording non-term units, recording multiple term forms and using TBX-Basic as a term record template.

7.5.1 *Sample and respondents' profile*

Given the specialized nature of the target population and the lack of formal marketing means available, we can consider that the call for participants gleaned a very satisfactory response. While the number of participants was slightly lower than in the first survey (159 vs. 168), thanks to a higher completion rate, it generated five more completed responses by participants who fulfilled all selection criteria (109).

Not only did the survey obtain a respectable number of responses, but the resulting sample had a good representation of both freelance and in-house translators, with an almost even split. As can be noted in the collection of industry surveys on this general topic listed in section 2.2.4 and described in Appendix B, in the translation industry, surveys often collect a higher percentage of freelance respondents. The higher number of in-house translators in this case may be due to the fact that many of our personal and professional acquaintances in the translation industry belong to the in-house sub-group and owing to the first-hand connection they may have more readily distributed the survey invitation to other contacts in the industry, likely also working in-house.

As was the case in the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, translators were also the predominate profession represented in the second survey. Not only was this result expected, given that in the translation industry translators are much more numerous than terminologists, project managers or revisers, it was also the desired distribution. The project's goal is to create guidelines for the design and creation of integrated termbases for the purpose of translation. Therefore, it was essential to obtain responses from translators with regard to the sub-hypotheses that required further testing. To obtain responses from this group, we trusted not only in the sheer power of numbers but also in the distribution of the invitation to individuals, associations and forums

whose membership and readership was composed mainly of translators.

The results obtained when we asked respondents about their profession brought to light an interesting evolution in the industry. Six respondents identified themselves as CAT tool specialists (5.5% of respondents). This indicates that since the advent of computer-aided translation tools in the early 1990s, their rise in popularity has given birth to a new job description in the translation industry.

The overview of the profile of our survey's respondents provides reassurance that the results come from an adequate sample that addresses our end goal.

7.5.2 Recording synonyms

The sub-hypothesis being tested in this question is the following: *Contrary to what current terminology and terminography literature recommends, translators working with TEnT-integrated termbases will organize their term records by equivalent pair rather than by concept.*

As explained in section 3.4, this sub-hypothesis is based on cases of synonymy where there is a direct relationship between source and target language equivalents, creating pairs. In such cases, if all source and target language equivalents are recorded on the same record, when either of the source language equivalents will appear in a text to translate all target language equivalents will be proposed, without any indication that one may be more suitable than the others (based on register, regional usage, etc.)

Based on the survey results, TEnT users working interactively prefer not to be presented immediately with the direct equivalent pair in order to have access to all possible equivalents for the term that appears in the source text. Both in the case of a term with synonyms that have two different registers and in the case of a term and its acronym, TEnT

users opted for the results that presented all forms and thus would have required the creation of a concept-based record. The response was stronger in the case of the acronym than in the regular term: 89% of respondents opted for the concept-based record when dealing with an abbreviated form, compared to the 74% of respondents who opted for the concept-based record in the case of the terms from different registers.

Although this result contradicts the initial sub-hypothesis, there is an important message in the user response. When working interactively, users want to have access to all the possible equivalents for a term. This is understandable for several reasons. Firstly, seeing the different equivalents of a term can help the translator to better understand the concept. For example, a translator may not be familiar with the scientific name of a drug but may be familiar with the commercial name. The scientific name will be preserved in the text but the translator would have gained a better understanding of the text being translated. Secondly, in cases of interchangeable synonyms, a translator may benefit from having the complete list of equivalents in order to use an alternate form to add variation to the text and thus avoid excessive repetition. Thirdly, translators may find it useful to be provided with equivalents of different language registers when a term match is found for translations that involve adapting the text to a different register (e.g. popularization of a specialized text, such as a drug description originally written for doctors that later needs to be distributed to patients).

The greater support for concept-based records in the case of acronyms is also understandable. TEnT users may prefer to always be presented with the acronym and its long form when translating as the same source language acronym may represent different concepts, e.g. IT stands for 91 different long forms according to the Acronym Finder

website⁶⁶ some of the most well-known ones are information technology, international trade or Italy. By being showed the full forms along with their acronyms, the translator may more easily understand the terms they represent and choose the target language acronym that corresponds to the concept being translated.

An additional question was added to the section on synonyms focusing on the case of a term in the source language that has multiple equivalents in the target language. Users were asked whether, when pretranslating, they would prefer to have all forms inserted in the target document or only the main form of the record (the first entered or the one identified as the main form, if the TEnT so allows). It was noted in the scenario that in order to see all forms inserted in the target text, all term equivalents would have to be entered as the main term entry, thus violating the principle of term autonomy.

The users' response was again to select the option that provides all forms of each term (64%), in spite of the resulting breach of the term autonomy principle. It seems as though users do not feel comfortable allowing the TEnT to make a decision for them or to always insert the same form for a term. Having all options inserted for each term allows the users to make an informed choice during the revision phase.

It seems clear that the user decision in each case was truly based on which results they find more useful when translating and not on what terminology management principles dictate. While for the interactive translation cases users opted for the concept-based record approach, as sanctioned by terminology management principles, in the pretranslation case they opted to record all forms of a term as the main entry, thus departing from the term autonomy principle. It is important to note that even in the group of eight terminologists, all but one agreed in this case to ignore the term autonomy principle in order to obtain the

⁶⁶ <http://www.acronymfinder.com>

more desirable results of having all forms of a term available in the pretranslated text.

For all their advantages, adopting these approaches would still have drawbacks. When translating interactively, the most logical equivalent will not always be presented as the first option, and when that happens, there will be a loss of time for the translator to select the second or third option as the equivalent to be inserted. When pretranslating, if all equivalents are entered as the main term entry in order to obtain all forms inserted in the target sentence, term autonomy will be lost. This can have an impact on the ease of manual searches if the termbase does not facilitate fuzzy or partial matches. In addition, it will also have an impact on the recording and display of supporting information for each of the forms. If all forms are displayed as one term, the context, note, and usage fields for each form would be all listed together under the main term and a referencing system would have to be devised to identify which form is linked to the supporting information. Finally, if the TEnT identifies the terms in the source document and the termbase by exact character-string matching, as most do, users will only be able to enter the multiple forms as the main term in the target language as otherwise no term match would be found for source terms. This would also render the termbase unidirectional.

From a user oriented-perspective, there is no other option but to decide which of the two scenarios is a lesser evil. In this case, TEnT users opted for concept-oriented records when translating interactively and the absence of term autonomy when pretranslating. Given the drawbacks described above, further research would be needed to determine whether the advantages outweigh the drawbacks or at the very least to find out if users continue to prefer concept-oriented records when the potential drawbacks are explained to them.

Tool vendors, however, could help to solve this dilemma by adding certain functionalities to TEnTs. Firstly, when the user is translating interactively and working with

concept-oriented records, the challenge of not being presented first with the most direct equivalent could be resolved if TEnT's added the functionality for establishing equivalent pair links within each record. This way, in the subject field of curling⁶⁷, a term record for the concept of *hammer*, which has the English and French equivalents of *last rock*, *last delivery*, *marteau*, *dernière pierre* and *dernier lancer* one could indicate *hammer* - *marteau*; *last rock* - *dernière pierre* and *last delivery* - *dernier lancer* as the logical synonym pairs. Ideally, this relationship could be indicated manually when creating a record from scratch but would be detected automatically when an equivalent pair is sent directly to the termbase from the TM or from the text being translated.

Secondly, in order to avoid the limitations that can derive from not preserving term autonomy in a record (e.g. silence when searching manually or pretranslating, unidirectional termbases), tool vendors could add an option for inserting all available target language equivalents when pretranslating, or not. If a user chose not to insert all target equivalents – and if the previous suggestion had been implemented – tool vendors could offer the possibility of inserting the direct equivalent pair.

7.5.3 Recording non-term units

The next sub-hypothesis to be tested was sub-hypothesis E, which as explained in section 3.5 focused on how receptive translators were to recording non-term units: *Contrary to what current terminology and terminography literature recommends, translators will record non-term units in their TEnT-integrated termbases.* The *Integrated Termbase Optimization Survey* tested this sub-

⁶⁷ Curling is a popular winter sport in Canada in which two teams compete to slide stones nearest to the centre of a target on an ice sheet.

hypothesis using non-term units such as email addresses, civic addresses, URLs or standard texts that change in the target language.

In the initial *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, respondents indicated that as a general rule they were not recording such units. The only case that showed some popularity – although it was by no means widespread – was the recording of common sentences, with 33% of respondents indicating they had term records containing such units. The rest of the non-term units were less prevalent if present at all: 16% of respondents recorded common paragraphs, 8% recorded URLs and 1% recorded emails.

However, when the question changed from whether they currently record such units to whether they would find it useful to do so in contexts where these require a change in the target language, the results were very different. In the *Integrated Termbases Optimization Survey*, respondents supported the creation of records for email addresses, civic addresses, URLs and standard texts. Each option obtained results ranging from 60 to 70% support.

The fact that even approximately two thirds of the small sub-group of terminologists supported the creation of records for email addresses, civic addresses and URLs shows that the benefits of recording such units are obvious to all types of TEnT users, who when working with integrated termbases may be open to recording non-term units.

Whether such units can be considered terms is a very complex question that falls beyond the scope of this project. However, they can certainly be considered translational units: specific expressions that recur in texts and that present a translation challenge.

It may be worthwhile recording these units for different reasons.

- a) These units can be long and they tend to be frequent; therefore having them recorded will spare the translator having to type them multiple times. Instead they can be inserted with a single click.
- b) These units tend to contain special characters (e.g. @, //, ~) or unusual capitalization that can be cumbersome to type. Inserting them directly from a termbase will be quicker.
- c) Any typographical errors in units such as email addresses, URLs or civic addresses can result in a complete malfunction of the hyperlink or a message being undeliverable. If the units are entered carefully and reviewed accordingly in a termbase, they can be easily inserted in target texts in the future with the peace of mind that no typographical error will make its way into that part of the text.
- d) There are typographical conventions for how civic addresses should be written (number of spaces between city and postal code, order of the street number, city, province/state, etc.). This may change from source to target language and from one client to the next, as they may adhere to different style guides. Correctly entering such units in a termbase would ensure that they are inserted following the required typographical norms.

When it came to recording standard text, respondents were strongly in favour of recording these units but the sub-group of terminologists was opposed to it.

While some users may argue that this type of unit belongs in the TM database rather than in the integrated termbase, the majority of respondents to this survey do see a value in recording *standard* text in integrated termbases, that is, not any text that may simply occur

more than once, but formulaic text that is meant to always be worded in exactly the same way (e.g. legal waivers, slogans, very formulaic document headers).

While these will be found in TM databases, TMs may contain multiple occurrences and being presented with the proper text to be inserted will depend on a number of factors. For example, it will depend on the tool having correctly aligned the corresponding segments. In addition, there is a risk that a previous translator may not have followed the preferred formulaic construction or may have introduced typographical errors or minor or major modifications to the text.

A well-maintained TM database will respond to the translator's need for such units. In particular, the segment-based TMs that allow for removal of identical segment repetitions and the inclusion of built-in maintenance tools to check for identical source language segments with different translations may be useful in this respect.

However, an integrated termbase will also serve the same purpose and it has a lower risk of containing multiple records for the same segment of text because creating a record for a specific unit requires a more conscious effort than sending a whole text to a TM.

7.5.4 Recording multiple term forms

Sub-hypothesis G stated that *[c]ontrary to what current terminology and terminography literature recommends, translators will record units in a TEnT-integrated termbase in all of their forms or their most frequent form(s).*

When looking at the general results, most respondents (59%) actually preferred recording only the base form. However, it must be also noted that 41% of respondents preferred to record multiple term forms. In order to increase the precision of the term matches, a sizeable portion of respondents were open to breaking with the terminographic

principle that requires that only the base form of each term be recorded.

As discussed in section 3.7, it is our opinion that recording multiple forms is more advantageous when translating with an integrated termbase. This is because, to the best of our knowledge, most TEnTs identify term matches based on character string matching, which means that as soon as a character in a term changes (due to any number of variations, including singular plural, masculine-feminine, present-past, etc.) no match will be found unless the tool offers a fuzzy match feature. Therefore, recording multiple forms of each term would ensure a higher number of matches.

There are some major potential drawbacks associated with implementing this approach. Firstly, there is the risk of investing excessive amounts of time recording multiple forms of a term that may not appear often in texts. This would result in a negative return on investment for those terms because the translator would have spent extra time recording unneeded terms. Secondly, having multiple forms for a concept may hinder manual searches in the database. When looking up an expression users will have to sift through all forms of each term. The number of results for each search may grow exponentially. Thirdly, if the termbase is shared or repurposed as a terminological reference work, having records with multiple forms of each term may be also a hindrance. For example, if users need to extract the content for a printed glossary, it may be difficult to select which form should be in the main entry. In addition, certain users such as technical authors or subject experts may judge the resource to be less reliable because it does not adhere to the principles of terminology.

The sub-hypothesis that it is beneficial for translators working with integrated termbases to part ways with the restriction of recording only base forms did not receive unanimous support in the survey, but it was popular among enough respondents to confirm the need to consider this option when designing a termbase and to carry out further research

on this topic.

When respondents were asked what forms they would record, only 9% opted for the most frequent form, 38% opted for all the forms found in the text they were translating and 53% actually indicated they would record all forms they could think of.

While having all possible source and target forms for each term in a termbase is definitely an ideal scenario when translating interactively, reaching this ideal scenario comes at a price: each form must be recorded and this can take time. As described above, the risk is as that time may be invested in recording forms that are very rare or that may never come up in the texts to be translated in the future. The main goal for a translator is to translate; there is limited time available to create records and therefore that time must provide a return on investment.

Therefore, it is important to note that this sub-hypothesis should not be read as a blanket rule dictating that all forms of a term should be recorded at all times. Rather it suggests that the base form may not always be the form that is most valuable to record. There are occasions when a few other forms may be worth recording alongside the base form or when a form other than the base form is the only one that is worth recording. We would highly recommend considering the frequency of forms and their complexity (e.g. irregular verb conjugations such as *beget, begat, begot, begotten, begetting*) when choosing which one to record. For example, in the case of a private corporation with only one *Chief Executive Officer*, there is obviously no need to record the plural form of the term. This strategy for optimizing a termbase can provide an immediate return in the form of more term matches and fewer instances of silence.

Tool vendors could contribute to facilitating the implementation of such a strategy and reducing the potential drawbacks described above. Some tool vendors have already tried to respond to the limited character string matching in tools by offering fuzzy term matching or stemmed term matching. Fuzzy term matching retrieves term records even when a certain number of characters differ between the term found in the source text and the term in the termbase. This type of matching will retrieve matches for different forms of a term (when the difference is of a small number of characters) but it may also retrieve items completely unrelated to the desired term (as sometimes the difference of a single character creates completely unrelated terms, as in *battle* - *bottle*, *thought* - *though*). Stemmed term matching retrieves term matches based on the stem⁶⁸ of each term. This approach is more likely to ensure that all term matches proposed will be related to the term found in the text to translate. However, the relationship to the term may not be a close one. If the relation is that of singular - plural, masculine - feminine, present - past, the proposed result is very likely to be useful, while if the relation is verb - noun (for example: *dictate* - *dictator*) the result is less likely to be useful.

Fuzzy and stemmed term matching are valuable features for improving term matches in a TEnT. However, it is important to understand that these features only allow the user to broaden the number of forms that will be linked to one term in the termbase. The equivalent that is proposed will remain in whichever form or forms were recorded by the user. Fuzzy and stemmed term matching increase the number of term matches but these matches must be adapted to the form actually found in the source text.

⁶⁸ In this context the *stem* of a term refers to the part of a word that can be found in all its inflected variants.

Tool vendors could further ease the implementation of the practice described in sub-hypothesis G in several ways.

- a) After identifying a record using fuzzy or term matching techniques, TEnTs could use linguistic parsing to identify the part of speech of the term in the source text and adapt the form of the proposed match to match that of the source text. This would save the translator a lot of time both at the recording stage (recording one form would be sufficient because multiple forms would be identified by the fuzzy or stemmed matching) and at the retrieval stage (because the proposed results would be adapted to the required part of speech). The drawback of this approach is that linguistic parsing is language-based and therefore it can be very costly to build in such a feature for all the languages supported by a tool. Alternatively, similar results could be obtained using example-based machine translation techniques that leverage the content of the TMs.
- b) When a new term is added to the termbase, the TEnT could do a statistical search in its TM databases for other forms of that term and their equivalents and propose them to the user to complete the record s/he is creating. This approach requires the intervention of the user who would have to validate the forms and equivalents proposed by the TEnT to be added to each record, but it would be easier than having to enter each form manually. As the results of such a feature would be based on statistics, the number of forms found and the quality of the equivalents would depend directly on the size of the TM available. In such contexts, the larger the TM, the more reliable the statistics would be. The benefit of working with statistics on a

user's texts is that the proposed results would reflect the frequency of each form in the specific texts that the user works with. Finally, this approach may be more appealing for tool vendors because it would be statistically based and therefore not limited to one language pair at a time.

- c) TEnTs could offer the possibility of easily identifying the base form of each term and creating a filtered view of the termbase displaying only this form or displaying it in a primary position. This would facilitate the use of the termbase for manual searches as well as for uses other than translation.

7.5.5 Using TBX-Basic as a term record template

Sub-hypothesis B proposed that [c]ontrary to the perceived desire for streamlining identified in sub-hypothesis a), translators will use a TBX-Basic-compatible term record structure if their TEnT has a built-in and modifiable template that follows this standard..

As described above nearly three quarters of our survey respondents agreed with the sub-hypothesis and indicated that they would use a built-in TBX-Basic compliant template to create a new termbase, if their TEnT offered that option. The positive support for this premise came from all subgroups: freelancers, in-house translators, respondents with or without formal terminology education, translators and terminologists. All groups gave a similar positive response ranging from 68.8% to 80% support. Therefore the proposed hypothesis is validated on the basis of the users' response.

A key element of the sub-hypothesis, however, lies in the TEnT providing a built-in TBX-Basic-compliant template. We take this opportunity to call on tool vendors to implement such a built-in template. Users' adoption of a TBX-compliant termbase template

will be advantageous not only to the users but also to the tool vendors.

As previously described, adopting a TBX-compliant termbase template ensures users on the one hand that their termbases will be easily exchangeable with any translation tool that adheres to this standard, and on the other hand, that they are using a termbase structure that has been validated by terminology experts and is internationally approved.

The benefit to tool vendors of users adopting a TBX-compliant termbase template is precisely that data exchange between tools will be smoother. The more users work with standardized termbase structures, the fewer cases of complicated data conversion projects tool vendors and users will have to face.

Users may need to exchange their terminological data with a colleague or client who uses a different tool or between their own tools when they have more than one. TEnT users have been reported to use several TEnTs (see section 5.5.1.2 or refer to LISA, 2004, p. 11 and Lagoudaki, 2006, p. 23).

Moreover, if termbase data exchange from one system to another is not an obstacle, users will be more open to evaluating TEnTs for their value as a whole and to switching tools if they find one that better meets their needs. Without this guarantee of exchangeability users may ultimately be locked in by the terminological resource they created and the impossibility of transferring to another tool. Some vendors may consider this a drawback, as it may allow users to easily move away from a tool. However, most tool vendors understand that customer retention is earned with the overall strengths of the tool and not by binding a customer to one system.

It could be argued that the built-in template in TEnTs is unnecessary because the structure is described in the standard. However, that suggestion would assume several things: first, that users will be aware that the standard exists; second, that they will easily find access

to the standard; and third, that they will be able to interpret the standard and all its requirements and structure their termbase accordingly. Certainly, many users would meet all of these conditions. However, others would not. If the template is provided by the tool vendor, it is far more likely that the standard will be applied correctly and that termbases will be truly exchangeable.

As welcome as the adoption of a TBX-Basic built-in template will be, it must be noted that it is not our intention to present TBX-Basic as the ultimate termbase structure that all members of the industry should adopt. TBX-Basic is a very well-thought-out structure that aims to meet the terminological needs of the translation and localization industry. Notwithstanding, it is a simplified version of TBX with a series of restrictions that may render it inadequate for the needs of some language service providers.

Its restrictions were the points of contention raised by those respondents who indicated they would not adopt TBX-Basic. Those who opposed TBX-Basic did so on the basis of either its simplicity or its complexity.

On one hand, certain users disliked the TBX-Basic structure because it lacked fields or field options that they consider to be essential for their termbase. This is very understandable: depending on the purpose of an integrated termbase, users may require a more complex structure. A termbase may have a double function; it may serve translation purposes and at the same time act as a reference tool for a wider audience. In large organizations, a termbase may be created to assist translators; however, complex terminology creation and approval processes may be in place that require other types of information to be recorded in addition to the basic fields proposed by the TBX-Basic template.

In such cases, the TBX-Basic template is simply not sufficient and a customized template will have to be designed. For such users, there is no quick and easy solution. The conceptual work of designing a template is unavoidable. Yet not all is lost for them because they could still obtain the desired exchangeability guarantee as long as they use a TEnT that supports TBX.

On the other hand, a number of respondents were opposed to the requirement that either POS, definition or context must be included in each record. These respondents wanted the freedom to be able to create extremely simple records with only source and target terms being mandatory.

Although it cannot be considered to be a universal solution, TBX-Basic can meet the needs of the majority of users (as the survey results indicate). One of the motives that inspired this research was the fact that new TEnT users faced the challenge of designing a termbase without much guidance available on how best to do so. The industry has made an effort to agree on a termbase exchange standard (TBX) and has provided a simplified structure for translation and localization purposes. While TBX-Basic will not be the ideal choice for every user, it can help users to get started on the right track, at least when it comes to the termbase design.

8 Integrated-termbase design guidelines

In this chapter we will present a series of guidelines on how to design and create an integrated termbase for the purpose of translation. These guidelines have been drawn from our previously presented hypothesis and sub-hypotheses as well as from general trends that the literature review or surveys have revealed. These guidelines are proposed not as a series of golden rules or as a perfect formula to design an infallible integrated termbase, but rather as a series of recommendations and most importantly a list of aspects to consider before a user embarks on the design and creation of an integrated termbase.

Any language professional knows that in this industry there is rarely a one-size-fits-all solution for anything, and terminology management within integrated termbases is no exception. Therefore, the following guidelines should be considered as general recommendations for TEnT users to evaluate in the light of their particular needs and context.

Moreover, it should be noted that the second user-acceptance test focused on *personal* integrated termbases in particular. Therefore, this fact must be kept in mind when designing a *shared* integrated termbase and when applying guidelines #7, #10, #11 and #12.

To begin, we propose the following overarching guideline:

#0 The termbase should be built not only as a stand-alone reference tool but also as a linguistic resource to be exploited within your TEnT. Therefore, the integrated termbase design and term recording strategies should be adapted to facilitate the automatic retrieval and insertion of term equivalents during the process of translating with a TEnT.

As explained in section 1.1.1, integrated termbases differ from standalone termbases because they serve as resources used by TEnTs to retrieve terms that appear in source documents, and to facilitate their insertion in the target text. Therefore, users of integrated termbases have specific needs that differ from users of term banks or other standalone termbases: how the TEnT retrieves term matches and proposes term equivalents can become an important factor in deciding how terms are recorded; the possibility of quickly inserting term equivalents in the target text may influence which units are recorded in the termbase; the fact that it is a translation resource limits the profile of the target audience and may impact which information is included on the record.

Terminology and terminography literature are useful resources for providing guidance on how best to build a termbase, but the different purposes of a standalone and integrated termbase must be kept in mind and terminology and terminography principles must be adapted, if necessary, to best serve the goals of the integrated termbase user.

The following more specific guidelines will help apply this overarching guideline to an integrated termbase. We have organized these guidelines following the chronological order in which decisions should be made when planning a termbase design and implementation: deciding the purpose, selecting and learning a tool, deciding on the number of databases, creating a record structure and recording guidelines, deciding on the details of a record structure and classification, selecting units to record, deciding how units will be recorded (e.g. synonyms, forms) and, once the database is in use, maintaining it. At the same time, these guidelines are also presented for the most part from the more general decisions (e.g. purpose, tool) to the more concrete (e.g. recording synonyms, forms) – with the exception of termbase maintenance, which is a general aspect but which is presented last

because of the chronological factor.

#1 The purpose of the termbase should be established from the start.

Although an integrated termbase by definition is found within a TEnT, it may have purposes other than being a terminological resource to assist translators while they work within a TEnT. The integrated termbase may be created with other applications in mind, such as providing a reference tool to establish proprietary terminology or specialized terminology. The target audience for this use may be the translation service but also the authoring, marketing, sales, customer or R&D services, who would usually look up terms manually.

Having a clear idea of the purpose(s) and end-user group(s) will help users to create a termbase design better suited to their needs.

It must be noted that this research project focused on integrated termbases for translators to be used within TEnTs and the overarching hypothesis that initiated this research is that when designing and creating an integrated termbase to be used for translation within a TEnT, users will adapt terminographic principles to this context. If the resulting termbase has purposes other than the one we just described, the recommended termbase design for such additional purposes must be researched and the impact of abiding by or disregarding recommendations for one or the other type of termbase must be evaluated.

#2 The terminology management system implemented should provide enough control over the termbases it will contain.

When creating an integrated termbase, users are often required to work with a software application that is already in place. However, if they have the luxury of being able to choose the software application to be implemented, it is important to select it carefully. The purpose(s) of a termbase as well as its design and term record template should be established before a software application is selected. The final solution should be flexible enough to manage the established termbase design and term record template.

If the termbase that is being created will have multiple purposes, the selected TMS should allow the creation of different views of the same content depending on the end-user (e.g. being able to block or display certain records or certain information in a record depending on which type of end-user is accessing the termbase). Such functionality already exists in certain TEnTs, such as the MultiTrans client-server environment with the Advanced Terminology Module.

#3 The terminology management system should be mastered in full.

In order to create and design an integrated termbase that best meets their needs, users must be aware of and fully master the different functionalities offered by the TMS.

As discussed in section 5.5.1.4, 70% of respondents to the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* had not mastered the advanced functionalities of their TMS. Given that the advanced functionalities are usually the ones that allow users to perfect, manage and maintain termbases, it is very important that termbase managers acquire the skills not merely of a regular user but of a super-user. Knowing how to import information into a termbase, customize the term record structure,

create and apply filters, identify and treat duplicate records or gather and analyze usage statistics will help to optimize a termbase. For example, depending on the functionality available in a TMS, if the proper options to create customized administrative fields and filters exist, a user may decide to store all term records in a single database because doing so will reduce the amount of maintenance required as there will be only one database to maintain rather than several.

#4 Whenever possible, terminology information should be centralized.

According to the results of the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey*, most respondents keep separate termbases mainly divided by subject and/or client. (For more details, see section 5.5.2.3).

As discussed in section 5.6.7, most textbooks on translation best practices and literature on terminology management recommend centralizing terminology in a single database to maximize the efficiencies of database maintenance as well as to avoid pigeonholing information.

This research project did not look into the reasons motivating users to keep separate databases, but a long discussion thread entitled *big 'mama' TDB* on Déjà Vu's online user forum Dejavu-L addressed this topic. The discussion concluded that if a user worked with clients in domains that did not overlap or if, when they overlapped, there were no terminology conflicts among the clients, a single termbase should be kept. If users worked with clients from domains that overlapped and whose terminology conflicted, multiple termbases were kept (ÓhAiseadha, 2001).

As the main reason to keep separate databases seems to be to avoid mixing clients' terminology, we would recommend keeping one centralized termbase as long as the TMS in use includes the possibility of creating customized administrative fields that allow for term records to be classified by client or domain, as well as filters that allow the user to exclude certain records from results based on the classification fields created. Should the TMS in use not provide these classification and filtering functionalities, and should a risk of using conflicting terminology for different clients exist, we would endorse keeping several termbases to avoid potential terminological mix ups.

#5 A basic term record structure and basic content selection and recording guidelines should be created before creating an integrated termbase.

No matter how the termbase is organized, which type of content is recorded or how, before starting to populate the termbase, users should consider why and how this resource will be used. Based on its purpose and future use, a basic set of rules on how to organize and enter information in the termbase should be created and followed. Section 5.5.2.2 illustrated how half of the respondents to the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* did not follow any basic guidelines. One of the aspirations of this research is to convince users of the added value of planning and to provide them with a simple instrument to assist them.

Part of the essence of a termbase is to store terminological information systematically and to facilitate the retrieval of information. In order to maximize these two goals of a TMS, it is important to abide by some basic rules to prevent the termbase from becoming a hodgepodge.

Planning the design and rules for content recording in a termbase does not need to be complex. A good start can be simply going through these recommendations and deciding if, and how, the user will follow each of them. Deciding not to follow one of these recommendations is already a step forward as the reader will have faced a particular design or recording issue, been made aware of its implications and decided a course of action.

#6 The number of fields on the term record structure should be limited to the essential information.

Translators do not need as many fields in a record structure as terminologists do. A traditional term bank record is created to assist any type of user to understand the term and its usage and to differentiate it from any other related concepts, while an integrated termbase record is created to avoid repeating the same terminological research multiple times or to facilitate insertion of the equivalent in a text.

Using a complex term record structure may deter users from filling out records and slow the creation of records as the user either has to enter a lot of information or select among multiple fields which is the most adequate for the information to be recorded in. Moreover, if most of the fields are optional and left empty, they can clutter the screen.

Selecting which fields are essential is a delicate task, as the fields to be included will vary from user to user depending on the particularities of their day-to-day work: language combination, types of documents translated and field of expertise. For example, a product code may be an essential piece of information for translators working in technical fields. Enough fields must be included to ensure that the term record can be reused in the future.

A good strategy is to always include a field for comments or observations in the structure. Such a field can be used to store miscellaneous information such as usage or grammatical notes, client preferences or cross-references. However, if users adopt this

strategy, they should monitor this field in case it would recurrently overflow with a hodgepodge of information. Should that happen, information recorded in this field should be classified and specific fields created, to facilitate the interpretation of the record as well as information lookup. The reason for this is that having to sift through half a page of unordered comments would be time-consuming and ineffective.

#7 Whenever possible, TBX-Basic should be used as a reference to create the term record structure.

Unless there is a need for a type of field or value that is not included in the TBX-Basic structure, it is recommended that users adopt this term record structure. Doing so spares the user having to establish a term record structure from scratch and instead provides him/her with an industry-approved standard structure. Moreover, as long as the TMS is TBX-compliant, using this record structure will facilitate the exchange of the termbase across TEnTs.

#8 Records should be classified not only by domain but also by client or project.

When translators work in multiple subject fields, classification by domain is very useful. However, translators will also benefit from keeping track of the client or project for which the term record was created. When keeping terminology centralized in a database, such classification can be achieved by means of administrative fields within the record. If the context for which the termbase is created requires keeping separate databases, such classification can be achieved by creating databases according to these criteria (e.g. by domain and client, domain and project or domain, client and project) or combining the use of separate databases and administrative fields to classify records within them (e.g. creating databases by client and domain, and classifying records within the database by project).

In centralized databases, these additional levels of classification, coupled with filtering and export functionalities in the TMS, allow users to create client- or project-specific glossaries to deliver along with their translations and more importantly, to monitor which client or project uses what term, and to identify and avoid terminology conflicts.

Some TMSs eliminate the additional effort required to fill out such fields for each record by allowing the user to set default values per session or permanently so that for any new record created a default client or project is entered automatically.

#9 Term records should contain enough information for future reuse.

Being able to reuse a term record in the future is the main motivation for recording terminological information. The more times a record is successfully looked up, the higher the return will be on the investment required to create the record.

Therefore, in an integrated termbase, users should record enough information for each term so that whenever it is looked up again, no further research is required. This recommendation can take many forms depending on each user's circumstances. In a personal termbase, if one opts for a minimalist approach, some records may only require the source and the target term for the user to be able to reuse them, while other terms will require a usage note or a definition. In a shared termbase, usually more information will be necessary (e.g. sources for the entry term, the equivalents or the definitions) so that other users can assess the reliability and appropriateness of the information found in the record.

As far as sources are concerned, although the respondents to the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* indicated that they mostly considered sources (of terms, contexts or definitions) to be optional or not included (see

more details on these results in section 5.5.4.1), we strongly recommend recording sources whenever possible. Even in a personal termbase for which the user is the author of the records, storing the sources of information not only allows the user to identify the reliability and appropriateness of the information but also makes the resource more easily shareable should the need arise. Advantages of doing so include the fact that we can reassess the quality of our records if we learn over time that a source is less reliable than we first thought or if a source becomes obsolete, or that if we ever share a large translation with other colleagues, they will be able to fully benefit from the termbase content.

#10 Units other than terms should be considered as candidate units to be recorded.

Terminographic principles dictate that the units to be recorded in termbases are terms. However, when using a TEnT to create an integrated termbase for translation, the termbase is used not only as a reference tool but also as a resource to facilitate the insertion of unit equivalents in the target text. Therefore, the units recorded do not need to be exclusively terms from specialized fields but can also be units (e.g. terms, phrases or any combination of words) that appear frequently in the texts to translate, units with a difficult spelling (containing special characters or simply a spelling that often results in typographical errors) or long units.

In short, users may wish to record any unit that appears frequently and that will be quicker to insert with a click or short-cut combination than to type from scratch or any unit that must be spelled in a specific way and for which they want to ensure no mistake is made. Phrases, collocations, standard text, email addresses, civic addresses, URLs or slogans have a place in termbases if they fall under the conditions described above.

#11 *The approach for recording synonyms should be adapted according to the translation method used.*

- A) *When creating a termbase for use in interactive translation, all synonyms for a concept should be recorded as terms on the same record.*

Adopting the concept-based record approach provides all the organizational benefits for which it was selected as an essential terminographic principle and more. When translating interactively, it also ensures that when one of the concept denominations appears in a text the translator will be presented with all equivalents available in the target language. Having access to all equivalents can help the translator to better understand the source language term, and it leaves the final decision of selecting the most adequate equivalent for each context in the hands of the translator.

While it is understood that in certain cases the logical relation between pairs of source and target synonyms of a term is very strong and can render synonyms not fully interchangeable in all contexts (e.g. due differences in the level of specialization or social register usage), recording synonyms by equivalent pair is not perceived as desirable by most respondents to the *Integrated Termbases Optimization Survey*. Instead most respondents preferred to be offered all options and to invest the time in identifying the equivalent to use in any given case.

To date, to the best of our knowledge, TEnTs do not offer a solution that would allow the identification of the logical equivalent pairs within a record while keeping all synonymous forms on the same record. We can only hope that this issue will be addressed in the future by tool providers.

- B) *When creating a termbase for use in pretranslation, all synonyms for a concept should be recorded as a single term on a record.*

When encountering a term with multiple target language synonyms during pretranslation, TEnTs will replace the term with one equivalent synonym – either the one marked as main target language entry or the first one created. In order to obtain all equivalents for a term and not allow the system to select the equivalent to be used, all equivalents must be entered as a single term.

The advantage of this approach to recording synonyms is that the translator will be able to select the desired equivalent among all options during the revision process and not be preconditioned by the selection made by the system.

It must be noted that breaking the term autonomy principle in such a way will make manual searches in the termbase more difficult (all searches will have to be of the type “the term field contains X” and not “the term field equals X”). Moreover, if the termbase will be used not only for pretranslation but also for interactive translation, this recording approach will hinder the insertion of terms when translating interactively as all options will be inserted when an equivalent is selected and all other equivalents will have to be deleted from the translated text. Finally, this strategy will also render the termbase unidirectional as explained in section 7.5.2.

At this time, when implementing these guidelines the user will have to decide which translation mode is more frequent in his/her own case and/or which of the challenges of the two approaches are ultimately likely to be less problematic, and opt for a synonym-recording approach accordingly. Perhaps in the future TEnT providers will offer the possibility of inserting all alternate equivalents when pretranslating, which would resolve this dilemma and would allow users to continue to preserve term autonomy.

#12 *In addition to the base form, the most frequent form(s) of a unit should also be recorded.*

When translating interactively or pretranslating, most TEnTs retrieve terms based exclusively on character-string matching, meaning that if the exact form of a term is not found in the termbase, no match is found. Therefore we recommend recording the base form and also the form(s) that appear frequently in texts to be translated.

The respondents of the *Integrated Termbases Optimization Survey* were divided when asked whether they would opt to record only the base form or multiple forms. While this guideline requires further testing to prove or disprove the benefits of recording multiple forms, we are in agreement with the 49% of respondents who supported recording multiple forms.

It must be noted, however, that recording multiple forms will not only increase the number of results obtained when translating interactively or pretranslating but also when carrying out manual searches. This is another example of a scenario where having a TEnT that allows the selection of the records or of the type of information on each record that will be displayed to different kinds of end-users would help implement an optimized integrated termbase without hindering the use of the termbase for different purposes (e.g. interactive translation vs. manual lookup). It must be noted that in such a case the base form should be systematically recorded so that a stand-alone termbase-like appearance can be provided to manual lookup users and an optimized integrated termbase that also contains inflected forms can be available to users translating interactively or pretranslating.

#13 TMs should be considered sources for locating and extracting term equivalents.

The taboo on using translated texts as references for terminology extraction is slowly disappearing for terminologists and has long been gone for translators. TMs are valuable corpora of aligned texts. If a TEnT includes functions that allow users to easily create term records from the content of the TMs, users should avail of it.

As with any other text, the key to using a translated text as a reference wisely is to be aware of the language quality and the reliability of its content. When using his/her own TMs, the user will likely know the context in which each translation was produced and the level and quality of terminological research carried out for each text. When using shared TMs, if there is any doubt about the accuracy of the equivalents used, additional verifications will be required before committing a unit to the termbase.

#14 Termbases should be maintained regularly.

This is good practice for any database, terminological or not. Although this guideline may appear self-evident, the negative consequences of disregarding it are devastating. An unmaintained termbase can quickly become cluttered, disorganized, outdated and incomplete. Such a termbase is consequently unreliable, which for a translator is the same as saying useless. Moreover, if integrated in a TEnT used for pretranslation, it can even become dangerous. For these reasons, we opted to include this task in our guidelines.

It is important to keep the records in a termbase up-to-date so that the information retrieved while translating remains relevant and pertinent. Corrections suggested by clients should be recorded, records created in a hurry should be reviewed and the termbase should be periodically assessed and cleaned up to eliminate duplicate records or obsolete information. If the TEnT allows for it, it can be useful to classify obsolete information under

a separate usage label (e.g. obsolete, avoid, refused) in order to keep the history of the term and for reference in case the term appears in texts to be translated in the future. A termbase will be only as good a resource as the information it contains.

8.1 Applying the guidelines

The above guidelines are presented as a series of aspects that a user should consider when setting out to design and implement termbase. We have presented them in a linear way and it would be easy to conclude that they must be implemented one after the other. However, it must be noted that in this project each guideline has been identified and validated independently. At this time, no test has been carried out on the combination of all guidelines.

However, the question of how these guidelines can or should be combined is a very legitimate one. Although no research has been carried out on this aspect, we can already observe that some guidelines will easily combine together and others will pose bigger obstacles.

A number of the guidelines can be followed independently from one another and therefore will be easily applicable, regardless of the context of use. These guidelines are the following:

- #1 The purpose of the termbase should be established from the start.*
- #2 The terminology management system implemented should provide enough control over the termbases it will contain.*
- #3 The terminology management system should be mastered in full.*
- #4 Whenever possible, terminology information should be centralized.*
- #5 A basic term record structure and basic content selection and recording guidelines should be created before starting an integrated termbase.*
- #8 Records should be classified not only by domain but also by client or project.*
- #9 Term records should contain enough information for future reuse.*
- #13 TMs should be considered sources for locating and extracting term equivalents.*
- #14 Termbases should be maintained regularly.*

As far as we can currently tell, the above guidelines do not present interrelated aspects that could cause a conflict. Guideline #5, which requires that a term record structure be created and guideline #9, which requires that the content be recorded in such a way that can be reused in the future are the only two that could potentially present a combination problem, but they are general enough for the two of them to co-exist without presenting an implementation problem. Finally, in the introduction of guideline #8, we have already discussed how it can be combined with guideline #4. We do not see a conflict between these two guidelines, only different forms of implementation.

This list leaves us with five guidelines that can be more problematic to combine. They are the following ones:

#6 The number of fields on the term record structure should be limited to the essential information.

#7 Whenever possible, TBX-Basic should be used as a reference to create the term record structure.

#10 Units other than terms should be considered as candidate units to be recorded.

#11 The approach for recording synonyms should be adapted according to the translation method used.

A) When creating a termbase for use in interactive translation, all synonyms for a concept should be recorded as terms on the same record.

B) When creating a termbase for use in pretranslation, all synonyms for a concept should be recorded as a single term in a record.

#12 In addition to the base form, the most frequent form(s) of a unit should also be recorded.

First, we can see a potential conflict between guidelines #6 and #7 in the case that the essential information for a user is less than the required information for the TBX-Basic structure. TBX-Basic requires the user to record source and target equivalents as well as either a definition, context or the unit's part of speech. The first survey revealed that none of the latter three fields were considered mandatory for the majority of respondents, so it is reasonable to expect that there will be cases where users do not find it necessary to record any of this supporting information. This could be because they are experienced translators who are very familiar with the terminology in their field and who use the termbase only as an automatic term lookup and insertion feature and not as a reference tool. It may also be the case that they find recording this information to be pointless because they work in fields in

which the terminology evolves at such a fast pace that records are created only for very short-term projects. In such scenarios, users may give more weight to guideline #6 and choose to disregard guideline #7.

Another case of potential contradiction in the combination of guidelines comes up when we consider the sum of guidelines #6, #7 and #10. As previously explained, TBX-Basic has been created as a simpler version of TBX, which was developed as a standard for traditional stand-alone termbases. Therefore, both standards aim to record information that describes terms. If we want to optimize our integrated termbase by recording non-term units such as URLs, civic addresses, telephone numbers, emails, phraseology while also using a TBX-Basic structure, we will still have to provide a definition, context or part of speech for these units. The easiest solution would be to provide a context. If we are translating a unit, it means we have a sample context to use. However, users may find it not essential to have a context, definition or part of speech for such units, which would go against guideline #6. To resolve this conflict, users will have to decide which guideline is more important for their own needs. Users will decide whether abiding by an internationally accepted standard, which ensures data exchangeability, is more important than the time that would be saved by not recording this non-essential information, or whether creating records efficiently is more important because the user can ensure his/her termbase exchangeability through other means (e.g. other formats shared both by colleagues and clients) or because exchangeability is not relevant. Moreover, some TEnTs can automatically insert the context when a term record is created. In such a case, the conflict would be significantly minimized as there would be no added time required to insert that information.

A third more complex combination is that of guidelines #11 and #12. Guideline #11 recommends recording all synonyms of a concept in the same record when translating interactively. However, if multiple forms of a term are recorded, there are some cases where users may prefer to keep forms that correspond directly in separate records. Examples might include storing plural forms on one record and singular forms on another or recording different verb tenses on different records. To give a concrete example, suppose we were to record the term *nuclear reactor* in an English-Spanish termbase for interactive translation. If we applied guideline #11 we would have a record including: *nuclear reactor, atomic reactor, reactor* in English and *reactor nuclear, reactor atómico, reactor de fisión, reactor* in Spanish. If we combined it with guideline #12, the record would now have double the number of terms: *nuclear reactor, atomic reactor, reactor, nuclear reactors, atomic reactors, reactors* in English and *reactor nuclear, reactor atómico, reactor de fisión, reactor, reactores nucleares, reactores atómicos, reactores de fisión, reactores* in Spanish. While users translating interactively preferred to see all the equivalents available for each term found, seeing the different number, gender and tense forms of each equivalent may be less useful. In most cases, singular units remain singular and verb tenses can be exchanged with some tenses in the target language, but not with all available tenses. Therefore, users will have to decide if in their particular case it may be more useful to record different forms of a unit's synonyms in separate records or if in such cases they prefer not to record multiple forms at all.

On the other hand, the same scenario must be considered in the case of pretranslation, for which guideline #11 recommends that all synonyms be recorded as a single term. If in addition to the synonyms we must add several available forms per synonym, the “equivalent” inserted may be several lines long. The translator will have to sift through double or triple the number of equivalents and delete all these alternate forms.

Again, users will have to decide if they prefer to separate different forms of a unit in different records or whether it is simply not effective to record multiple forms for terms with synonymous equivalents. The combination of these guidelines would certainly need further research.

Finally, a case of conflict that has been already discussed is that of termbases that have more than one purpose (e.g. an integrated termbase for translation that must also serve as a terminology reference tool for technical writers and customer support representatives). According to the guidelines, in order to optimize the termbase for translation within the TEnT we should classify records by client or project, record non-term units and if the termbase is used for pretranslation, we should record all synonyms as a single term and record multiple forms of each term. These strategies to optimize the termbase for translation can render it unusable for other groups of users. In this case the solution would be to add a feature to the termbase that makes it possible to block certain content (records, units, fields) according to the user profile. However, if guideline #11B is applied and all term equivalents are entered as a single term not even the type of granular access-right control as described above would render those records acceptable for other audiences. In such a case, if the translation service typically pretranslates their texts and cannot do without inserting all alternate forms of a term, then it may be preferable to disregard guideline #4 and keep two separate termbases—one for reference and one for pretranslation—to benefit a wider group of users. This could also be the preferred solution if the TMS in question does not offer the possibility of setting up such sophisticated content blocks based on the user profile.

These are the difficulties we can foresee at this stage when trying to combine the guidelines. There will certainly be more depending on each particular scenario. As mentioned at the beginning of this chapter, because each translation service is likely to have some

unique needs, each user should weigh the guidelines according to their own needs. On our part, we recognize that additional research is required to be able to provide improved recommendations on how to combine the guidelines in different contexts. With the present research we have simply laid a foundation upon which we intend to build more solid principles in the future.

9 Concluding remarks

The concluding chapter of this thesis looks both backward and forward. We begin with a brief assessment of the initial objectives, followed by a reflection on the work accomplished and an evaluation of the potential impact of the results of this research. Next, we will look to the future and consider the paths that this research may take going forward. We may have reached the end of this research project, but research never ends.

9.1 Assessment of the objectives

This project outlined the following objectives in section 1.2: a) to identify the current practices in terminology management for translation within TEnTs, b) to establish a series of user-controlled strategies to optimize integrated termbases for translation within this same type of environment and finally c) to evaluate these user-controlled strategies based on user acceptance.

We are confident that each of these objectives has been successfully met. We presented a snapshot of current terminology management practices in chapter 5. Not only did we discuss the findings of the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* but we also contextualized them with reference to a collection of recent surveys on translation memory system use and terminology as well as to the existing literature on this subject matter. Based on the literature review, personal experience and observations on the current practices in terminology management with TEnTs, we were able to formulate a series of user-controlled strategies to optimize integrated termbases. These strategies were first introduced in section 1.3 and in chapter 3 and they were later refined in chapter 6 based on the data collected on current practices.

Finally, they were transformed in guidelines in chapter 8, after having been tested for their user-acceptability. This last step represented the completion of the third objective which necessitated asking users to provide their opinion on the proposed user-controlled strategies and to assess whether they found them useful for optimizing integrated termbases. Chapter 7 provided an account of the user acceptance test and the feedback collected from users and analyzed the results obtained.

Although we feel this project has met its objectives, we know the work is not finished. As the project evolved and took shape we identified a number of possible ramifications and additional limitations that were not identified at the onset. The following section will put the project's success in context.

9.2 Lessons learned

Looking back at this research project, overall we are very satisfied with what we have accomplished. The two surveys met their objectives of first identifying the current practices on terminology management within TEnT's and later testing the acceptance by users of a selection of the sub-hypotheses of this research project.

The number of answers collected exceeded expectations in both cases. In fact, the positive response to the first survey motivated a redefinition of the focus of this project. What was originally intended to be a preliminary overview turned out to represent a collection of data which warranted an in-depth study.

The combination of results obtained during the *Use of Terminology Management Systems Integrated with Translation Environment Tools Survey* and the *Integrated Termbases Optimization Survey* coupled with the literature review and our practical experience constituted a solid basis for

the formulation of the basic guidelines presented in chapter 8.

However, the lack of empirical testing of this project's hypothesis and sub-hypotheses prevent the resulting guidelines from being presented as a set of rules or principles. At this stage, the guidelines must remain in the form of recommendations based on subject-matter literature, current practices and user acceptance.

Moreover, due to the variety of scenarios in which translation takes place (bilingual and multilingual settings, in-house translation services, freelance practice, localization services) the above guidelines can certainly not be considered universal.

Furthermore, as discussed in section 8.1, the guidelines have been established and tested in isolation and not in combination with one another. Further research is needed to establish the positive or negative impacts and ramifications of combining all or some of the guidelines presented in this research.

Another limitation to be taken into account is that these guidelines cannot be considered to be timeless, given the speed at which TEnT's evolve. TEnT providers are constantly improving their tools and we certainly hope that these improvements will translate into functionalities that will facilitate the recording and retrieval of terms. Thus, some of the guidelines above will most likely become obsolete over time.

The exception to these limitations may lie with the overarching guideline. The recommendation to be aware of the inherent differences between stand-alone and integrated termbases and to adapt the integrated termbase design and term recording strategies to these particularities shall remain valid for all translation scenarios and times. What will change is the how.

9.3 Call to the community

In spite of this project's limitations, we believe that the different members of the translation community can benefit from what has been accomplished. The first and most direct group of beneficiaries are without a doubt the translators. This work represents a step toward filling the gap in the literature on terminology management within TEnTs. We hope that translators will be able to turn to our work and find guidance when facing the challenge of designing and implementing an integrated termbase. Not only that, but we also believe that this work contributes to reinforcing the importance of a) the role of terminology in translation, b) the unique features of termbases integrated with TEnTs (automatic term lookup and insertion) and c) the optimization of integrated termbases to best take advantage of such features. What we most want translators to take away from this thesis is the idea that they are not only entitled but welcome to adapt their integrated termbases to best suit their needs within a TEnT, be it by applying the set of guidelines proposed herein or by creating their own optimization strategies.

Second, the description of current practices, the set of guidelines, the discussions on the needs of translators working with TEnTs, as well as the description of the advantages, drawbacks and side-effects of implementing such guidelines, all provide terminologists with additional insight into the particular needs of translators working with TEnTs. Any terminological work is always carried out with an audience and a purpose in mind. In addition, whether working in large organizations or as freelancers, terminologists are often in charge of creating and managing termbases for translators working with TEnTs. We hope terminologists regard this thesis not as an attack against terminology principles but as an attempt to explore the need for an additional type of terminological product and the ways in which it could be achieved. We would be honoured if terminologists saw in these guidelines

seeds that could lead to better integrated termbases. Seeds that terminologists, thanks to their methodological background and their often pivotal positions in managing termbases for translators, could cultivate and grow in their own products.

Third, this thesis provides translator trainers with an additional tool for disseminating information about the importance of terminology and the differences between integrated termbases and stand-alone ones. It also provides a basic recipe for a reflection on how best to design and implement an integrated termbase. The eCoLoTrain survey (2006) revealed that a certain number of translator trainers did not feel comfortable providing their students with training on TEnTs and TMSs. We believe that the paucity of literature on this topic contributed to this lack of confidence and we hope that this project provides trainers with additional documentation to help them strengthen their curricula with regard to the use of TEnTs and TMSs for translation.

Fourth, this thesis is intended to partially fill a gap in translation technology literature but also – and more importantly – to be a catalyst for discussion, reflection and investigation on terminology management within TEnTs. As we will summarize in the next section, our work on this thesis has left us with seemingly endless questions and possibilities for further research. The ultimate success for this thesis would be not only to see the guidelines implemented and embraced by the industry, but also to see other academics take on the challenge of pursuing, deepening, expanding or challenging the idea of adapting termbase design and implementation for translation-oriented termbases integrated with TEnTs.

Last but not least, for translation software developers, this thesis can serve as a snapshot of how integrated termbases are used on a daily basis as well as a repertoire of scenarios upon which current TMSs can improve. Throughout this project we have proposed functionality that could ease the implementation of these guidelines or even

eliminate their need. Some of the ideas that we have brought to the table are: the inclusion in TEnTs of a TBX-Basic built-in-record template to facilitate the creation of termbases (see section 3.2), the possibility of establishing links between equivalent synonym pairs within records (see section 7.5.2), the option to insert all alternate forms or only the base form in the target text when pretranslating (also in section 7.5.2) or the functionality to suggest to the user different forms of a same unit, either for insertion in the target text or for inclusion in a record, based on linguistic parsing or example-based machine translation techniques using TM content (see section 7.5.3). Without a doubt, translation software developers' skills and creativity can not only bring such ideas to fruition, but also certainly find countless additional improvements and solutions to address the particular needs of translators working with TEnTs as described in this thesis. Ironically, we would not regret seeing this thesis become obsolete as a result of translation software developers finding automated solutions to the challenges that these guidelines attempt to resolve.

9.4 Future research

A plethora of research avenues certainly stem from this project. In particular, we would like to carry out empirical, quantitative and more targeted tests based on different translation profiles.

Firstly, empirical research could aim to uncover the motivation behind the answers provided by users. The surveys in this research project collected information about a series of current practices as well as user judgements on proposed sub-hypotheses. In both cases, the data was collected in the abstract. In the first survey, the respondents' perception of their own practices may not correspond directly to reality. Meanwhile, in the second survey, participants may react differently to a strategy presented in a concrete scenario (as in the

survey) than they would if asked to apply that approach on a daily basis. Future research could explore the reasons why translators opted to classify their records by project as opposed to client, to not record sources for a term, context or definition, to choose to record the base form instead of multiple forms of a term and so on. Such research could be carried out through interviewer-administered surveys, as discussed in section 4.1. The results of such research would shed light on the different factors influencing a user's decisions on how a termbase is designed and a term recorded, which in turn would make it possible to adapt the guidelines to better reflect the different possible translation scenarios.

Secondly, the research would focus on testing the impact of each sub-hypothesis and the combination of some or all of them on productivity, quality and consistency. To test productivity we could measure the time required to translate a given text with or without an optimized integrated termbase. This research could be carried out in the form of case studies or experiments using integrated termbases in which one or all guidelines had been applied. To test quality we could measure the number of errors prevented by the use of an optimized integrated termbase. In this instance, a case study would be the most conducive method. An example of a test could be as follows. A translation service initially using a termbase that was created without applying any of the guidelines could be studied during a set period of time in order to obtain a baseline number of typing errors, term equivalent misuse or inconsistency errors for which a term record existed and therefore should have been preventable. Their integrated termbase would then be modified to apply the suggested guidelines. The service would work for another set period of time with the newly optimized termbase and the same measurement of preventable errors would take place. If the guidelines are indeed useful for improving quality, one would expect that the number of preventable errors should decrease in the second period of testing. Finally, to test consistency we could measure the number of

unwanted terminological inconsistencies found in a set of translations carried out with a non-optimized integrated termbase vs. an optimized one. For such a test, an experiment could be set up. However, it would probably be logistically more feasible to carry out a case study similar to the one described above or even to test both factors in combination.

Lastly, given the variety of contexts in which translation takes place (e.g. in-house translation service, freelancing, or localization service) it would be interesting to test the proposed guidelines in different work settings and for different language pairs. A combination of the above qualitative and quantitative tests could be carried out to establish whether the guidelines apply equally to different work settings, language combinations, domains and tools or if they vary, and if so, how.

9.5 Closing remarks

As presented in section 2.1, there is an increasing interest in and awareness of the relevance of managing terminology in the language industry. Managers are sure to pay attention to arguments such as “[...] investment in terminology standardization before problems arise avoids a ninefold multiplication of costs later” (Childress, 2007, p. 44). Clearly, faced with this argument it is evident that a company cannot afford not to manage terminology.

If companies cannot afford not to manage terminology, can they afford to invest in terminology management and without consistently retrieving term matches or optimizing integrated termbases?

10 Bibliography

- AAPOR. (American Association for Public Opinion Research). (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Retrieved on April 22, 2012 from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156
- Aday, L.A. (1996). *Designing and Conducting Health Surveys: A Comprehensive Health Guide*. 2nd Ed. San Francisco: Jossey-Bass.
- AILIA (Language Industry Association). (2004). *The Canadian Translation Industry: Technology Roadmap*. Retrieved March 28, 2009 from http://www.crtl.ca/docs/The_Canadian_translation_industry.pdf
- ALPAC (Automatic Language Processing Advisory Committee). (1966). *Languages and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council*. Retrieved March 28, 2009 from http://www.nap.edu/catalog.php?record_id=9547#toc
- Arrouart, C.; Bédard, C. (2001). “In Praise of the Bitext”, *Circuit*, pp. 73, 30. Retrieved February 19, 2008 from http://www.terminotix.com/eng/info/mem_5.htm
- Arthern, P. (1979). Machine Translation and Computerized Terminology Systems. A Translator’s Viewpoint. In B.M. Snell (Ed.), *Translating and the Computer. Proceedings of a Seminar. London, 14th November, 1978*. Amsterdam/New York/Oxford: North-Holland Publishing Company.

- ATIO (Association of Translators and Interpreters of Ontario). (2007). *Results of the 2007 Survey of Salaried Translators*. Association of Translators and Interpreters of Ontario. Retrieved September 6, 2009 from http://www.atio.on.ca/Membership/Sal_Survey/Sal_Tran_Srvy_Rslts.asp
- ATRIL (2003). *Déjà Vu X Standard Users' Guide*. Retrieved December 15, 2009 from <http://www.atril.com/docs/DVX/DVX%20Standard.pdf>
- Auger, P. (1989). "Informatique et terminologie: revue des techniques nouvelles". *Méta*, 34(3), pp. 485-492.
- Austermühl, F. (2001) *Electronic Tools for Translators*. Manchester: St. Jerome Publishing.
- Balch, C.V. (2010). *Internet Survey Methodology*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Ballard, M. (2002). *La traduction de l'anglais au français. Pour s'initier à la pratique de la traduction de façon efficace et rigoureuse*. Malesherbes (France): Nathan.
- Beidernikl, G.; Kerschbaumer, A. (2011). "Sampling in Online Surveys". In A. Rodney, R. Woods and J.D. Baker(Eds.), *Handbook of Research on Electronic Surveys and Measurements*, Hershey/London: Idea Group Reference, pp.90-96.
- Benis, M. (1999). "Translation from O to R", *TransRef*. First published in *ITI Bulletin*. Retrieved May 4, 2008 from <http://www.transref.org/default.asp?docsrc=/u-articles/Benis3.asp>
- Bergeron, M. (2001). "big 'mama' TDB", *Dejavu-L Yahoo Group*, May 28. Retrieved December 14, 2009 from <http://tech.groups.yahoo.com/group/dejavu-l/message/11993>
- Bethlehem, J.; Biffagnandi, S. (2012). *Handbook of Web Surveys*. Hoboken, NJ: John Wiley & Sons.

- Bourigault, D.; Slodzian, M. (1999). "Pour une terminologie textuelle", *Terminologies Nouvelles*, 19, pp. 29-32.
- Bourque, L.B.; Fielder, E. P. (2003). *How to Conduct Self-Administered and Mail Surveys*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- Bowker, L. (2002a). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- Bowker, L. (2002b). "An empirical investigation of the terminology profession in Canada in the 21st century", *Terminology* 8(2), pp. 283-308.
- Bowker, L. (2004). "What Does It Take to Work in the Translation Profession in Canada in the 21st Century? Exploring a Database of Job Advertisements", *Meta*, XLIX(4), pp. 60-972.
- Bowker, L. (2005). "Productivity vs. Quality. A pilot study on the impact of TMs", *Localisation Focus* 2005-2006, pp. 133-140.
- Bowker, L. (2011). "Off the record and on the fly: Examining the impact of corpora on terminographic practice in the context of translation". In A. Kruger, K. Wallmach and J. Munday (Eds.), *Corpus-based Translation Studies: Research and Applications*. London/New York: Continuum, pp. 211-236.
- Bowker, L.; Marshman, E. (2009). "Better integration for better preparation: Bringing terminology and technology more fully into translator training using the CERTT approach". In *Terminology* 15(1), pp. 60-87.
- Bowker, L.; Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*, London/New York: Routledge.
- Brigham Young University. (2009). *Graduation: Statistical Highlights*. Retrieved November 12, 2011 from <http://saas.byu.edu/registrar/graduation/statistics.php>

- Brown, M. K. (2003). "Trends in Writing for Translation." *Multilingual Computing and Technology* 14 (7), Special Supplement #59 on Writing for Translation, pp. 4 – 8.
- Cabré, M.T. (1992). *La terminologia: La teoria, els mètodes, les aplicacions*. Barcelona: Empúries.
- Cabré, M.T. (1998). *La terminologie: Théorie, méthode et applications*. Ottawa: Presses de l'Université d'Ottawa.
- Cabré, M.T. (1999). *La terminología: Representación y comunicación*. Barcelona: IULA.
- Champagne, G. (2004a). *Portrait of Terminology in Canada*. Unpublished report submitted to the Translation Bureau of Canada, Public Works and Government Services Canada, 38 pp.
- Champagne, G. (2004b). *The Economic Value of Terminology: An Exploratory Study*. Unpublished report submitted to the Translation Bureau of Canada, Public Works and Government Services Canada, 36 pp.
- Childress, M.D. (2007). "Terminology work saves more than it costs". *Multilingual*, April/May 2007, pp. 43-46.
- Cronin, M. (2003). *Translation and Globalization*. London/New York: Routledge.
- CTISC (Canadian Translation Industry Sectoral Committee). (1999). *Survey of the Canadian Translation Industry: Final Report of the Canadian Translation Industry Sectoral Committee, September 1999*. Retrieved October 19, 2009 from <http://www.uottawa.ca/associations/csict/represum.pdf>
- Delisle, J. (2003). *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*, 2nd edition reviewed and expanded. Ottawa: University of Ottawa Press.

- Dillon, S.; Fraser, J. (2006). "Translators and TM: An Investigation of 'Translators' Perceptions of Translation Memory Adoption." *Machine Translation* 20(2), pp. 67-79.
- Dubuc, R. (2002). *Manuel pratique de terminologie*, 4th edition. Montreal: Linguattech.
- Dunne, K.J. (2007). "Terminology: ignore it at your peril". *Multilingual*, April/May 2007, pp. 32-38.
- Durán Muñoz, I. (2010). "Specialised lexicographical resources: a survey of translators' needs". In S. Granger y M. Paquot (Eds.), *eLexicography in the 21st century: New Challenges, new applications. Proceedings of ELEX2009. Cahiers du Cental*, vol. 7. Louvain-la-Neuve : Presses Universitaires de Louvain. pp. 55-66. Retrieved on December 11, 2011 from http://uma.academia.edu/IsabelDuran/Papers/800731/Translators_Needs_into_Account_A_Survey_on_Specialised_Lexicographical_Resources
- EAGLES. (1996). *Evaluation of Natural Language Processing Systems. Final Report*. Retrieved February 19, 2008 from <http://www.issco.unige.ch/projects/ewg96/>
- eCoLoTrain. (2006). "eCoLoTrain Results. 'Translator Training Survey.'" *Institute of Translation & Interpreting*, UK. Retrieved August 29, 2009 from <http://www.iti.org.uk/uploadedFiles/surveys/eCoLoTrain-Results%20April%202006%20graphic.pdf>
- Esselink, B. (2000). *A Practical Guide to Localization*. Amsterdam/Philadelphia: John Benjamins.
- Estopà, R. (2001). "Les unités de signification spécilisées: élargissant l'objet du travail en terminologie", *Terminology* 7(2), pp. 127-237.
- Fidura, C. (2007). "The benefits of managing terminology with tools". *Multilingual*, April/May 2007, pp. 39-41.

- Fulford, H.; Granell-Zafra, J. (2005). "Translation and Technology: A Study of UK Freelance Translators." *The Journal of Specialized Translation* 4 (July), pp. 2-17. Retrieved September 6, 2009 from http://www.jostrans.org/issue04/art_fulford_zafra.pdf
- Gaudin, F. (1993). *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Presses de l'Université de Rouen.
- Gauthier, F. (2006). *Survey on Rates and Salaries*. Trans. James Cookson. Ordre des traducteurs, terminologues et interprètes agréés du Québec. [Available to members only].
- Gauthier, F. (2008). *Sondage de 2008 sur la tarification et les salaires. Résultats Destinés aux membres*. Ordre des traducteurs, terminologues et interprètes agréés du Québec. [Available to members only].
- García, I. (2010). "Beyond Translation Memory: Computers and the Professional Translator", *JoSTrans: The Journal of Specialized Translation*, Issue 12. Retrieved on December 4, 2011 from http://www.jostrans.org/issue12/art_garcia.php
- Gow, F. (2007) "You Must Remember This: The Copyright Conundrum of "Translation Memory Databases", *Canadian Journal of Law and Technology*, 6(3), pp. 175-192.
- Guppy, N.; Gray, G. (2008). *Successful Surveys. Research Methods and Practice*, 4th ed. Toronto: Thomson Nelson
- Heyn, M. (1998). "TMs: Insight and Prospects". In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in Diversity? Current Trends in Translation Studies*, Manchester: St. Jerome Publishing, pp. 123-136.
- Höcker, M. (2003) *Translation Memory Survey 2003*. eContent Localization Resources for Translator Training, Retrieved September 6, 2009 from http://ecolore.leeds.ac.uk/downloads/2003.05_bdue_survey_analysis.doc

- Holmes, J.S. (1972). “The Name and Nature of Translation Studies”, expanded version in *Translated! Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi, 1988, pp. 66-80.
- Hutchins, J. (1998). The origins of the translator’s workstation, *Machine Translation*, 13(4) pp. 287-307. Retrieved February 19, 2008 from <http://www.hutchinsweb.me.uk/MTJ-1998.pdf>
- ISO (2000). “ISO 1087-2:2000, Terminology work – Vocabulary – Part 2: Computer applications”, *International Organization for Standardization*. Retrieved December 15, 2009 from http://www.iso.org/iso/catalogue_detail.htm?csnumber=32819
- ISO (2008). “ISO 30042:2008 Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX)”, *International Organization for Standardization*. Retrieved December 15, 2009 from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45797
- ISO (2009). “ISO 12620:2009 Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources”, *International Organization for Standardization*. Retrieved December 15, 2009 from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37243

- ISO (2009). “ISO/DIS 26162 Systems to manage terminology, knowledge and content – Design, implementation and maintenance of terminology management systems”, Draft, *International Organization for Standardization*, Retrieved March 3rd, 2010 from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43427
- ITI (Institute of Translation & Interpreting). (2007). *Translation Memory Survey*. Retrieved August 29, 2009 from <http://www.iti.org.uk/uploadedFiles/surveys/TMSurvey-Results.pdf>
- Jaekel, G. (2000). “Terminology Management at Ericsson”. In R. Sprung (Ed.), *Translating into Success: Cutting-Edge Strategies for Going Multilingual in a Global Age*, Amsterdam/Philadelphia: John Benjamins, pp. 159 – 171.
- Joscelyne, A. (2000). “The Role of Translation in an International Organization”. In R. Sprung (Ed.), *Translating into Success: Cutting-Edge Strategies for Going Multilingual in a Global Age*, Amsterdam/Philadelphia: John Benjamins, pp. 81 – 95.
- Kay, M. (1980[1997]). The proper place of man and machines in language translation. Reprinted in *Machine Translation* 12(1/2), pp. 3-23.
- Kelly, N.; DePalma, D.A. (2009). *The Case for Terminology Management. Why Organizing Meaning Makes Good Business Sense*. Lowell: Common Sense Advisory. Retrieved September 3, 2009 from http://www.commonsenseadvisory.com/research/report_view.php?id=99&cid=0
- Kenny, D. (1999). “CAT Tools in an Academic Environment: What are They Good for?,” *Target : International Journal of Translation Studies*, 11(1), pp. 65-82.
- L’Homme, M.-C. (2004). *La terminologie: principes et techniques*. Montreal: Linguattech.
- L’Homme, M.-C. (2008). *Initiation à la traductique*, 2e éd.. Montreal: Linguattech.

- Lagoudaki, E. (2006). "TMs Survey 2006: Users' perceptions around TM use". In proceedings of the *ASLIB International Conference Translating & the Computer 28*, London, UK, 15-16 November. Retrieved February 19, 2008 from http://www.lagoudaki.com/research/TMSurvey_2006_Imperial_College_London.pdf
- Lanctôt, F. (2001). "Splendeurs et petites misères...des mémoires de traduction". *Circuit*, pp. 72, 30. Retrieved November 20, 2007 from http://www.terminotix.com/t_fr/info/mem_4.htm
- LISA (Localization Industry Standards Association). (2007). *The Globalization Industry Primer. An introduction to preparing your business and products for success in international markets*. Retrieved September 6, 2009 from <http://www.lisa.org/Globalization-indust.468.0.html>
- LISA (Localization Industry Standards Association). (2008). *Terminology Special Interest Group*. Retrieved January 13, 2009 from <http://www.lisa.org/Terminology-Special.102.0.html>.
- LISA (Localization Industry Standards Association). (2009). *TBX-BASIC*. Retrieved March 14, 2010 from <http://www.lisa.org/TBX-Basic.926.0.html>
- Lommel, A. (Ed). (2002). "LISA Translation Memory Survey: Translation Memory and Translation Memory Standards". *Localization Industry Standards Association*. Retrieved September 6, 2009 from <http://www.lisa.org/Translation-Memory-S.523.0.html>
- Lommel, A. (2004). "LISA Translation Memory Survey: Translation Memory and Translation Memory Standards", Ed. Rebecca Ray. *Localization Industry Standards Association*. Retrieved September 6, 2009 from <http://www.lisa.org/Translation-Memory-S.518.0.html>

- Lommel, A. (Ed.). (2005). "LISA Terminology Management Survey: Terminology Management Practices and Trends", *Localization Industry Standards Association*. Retrieved January 13, 2009 from <http://www.lisa.org/Terminology-Management.461.0.html>.
- Manfreda, K.L.; Berzelak, N.; Vehovar, V. (2011). "Nonresponse in Web Surveys." In Lobric, M., Ed. *International Encyclopedia of Statistical Science*. Berlin/Heidelberg: Springer, pp. 984-987.
- Marshman, E. and Bowker, L. (in press). "Translation technologies as seen through the eyes of educators and students: Harmonizing views with the help of a centralized teaching and learning resource". In Hubscher-Davidson, S. and Borodo, M. (Ed.), *Global Trends in Translator and Interpreter Training*, London/New York: Continuum.
- Martí Antonín, M.A.; Alonso Martín, J.A.; Badia Cardús, T.; Campàs Montaner, J.; Gómez Guinovart, X.G.; Gonzalo Arroyo, J.; Llisterri Boix, J.; Rafael i Fontanals, J.; Rodríguez Hontoria, H.; Soler i Bou, J.; Verdejo Maíllo, M.F.. (2003). *Tecnologías del lenguaje*. Barcelona: Editorial UOC.
- McBride, C. (2009). *Translation Memory Systems: An Analysis of Translators' Attitudes and Opinions*. M.A. Thesis. University of Ottawa: Canada
- McInnis, N. (2008). "On the Lighter Side: Terminology Results," *InformATIO* 37(4), p. 4.
- McInnis, N.; Takla, M. (2005). "Survey of Independent Translators", *Association of Translators and Interpreters of Ontario*. Retrieved September 6, 2009 from http://www.atio.on.ca/Membership/Ind_Survey/Survey05_Intro.html.

- Melby, A.K.; Smith M.R.; Peterson, J. (1980). "TTS: interactive translation system", *Coling 80: Proceedings of the Eighth International Conference on Computational Linguistics*, Tokyo, Japan, pp. 424-429. Retrieved February 19, 2008 from <http://acl.ldc.upenn.edu/C/C80/C80-1064.pdf>.
- Mel'čuk, I. (2004). "Actants in semantics and syntaxI: actants in semantics", *Linguistics*, 42(1), pp. 1–66. Retrieved on December 11, 2011 from http://www.coli.uni-saarland.de/~tania/CMGD/Actants_in_semantics.pdf
- MultiCorpora. (2010). *What's New MultiTrans V4.4 R2*. Retrieved November 6, 2011 from http://www.multicorpora.com/filesNVIAdmin/File/WhatsNew44R2_LowRes_EN.pdf
- MultiLingual. (2007). *Terminology*, April-May.
- Newman, I.; McNeil, K. (1998). *Conducting Survey Research in the Social Sciences*. Lanham/New York/Oxford: University Press of America.
- Newmark, P. (1988). *A TextBook of Translation*. New York/London/Toronto/Sydney/Tokyo: Prentice Hall.
- O'Brien, S. (1998). "Practical Experience of Computer-Aided Translation Tools in the Software Localization Industry". In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in diversity? : Current trends in translation studies*. Manchester: St. Jerome, pp. 115-122.
- ÓhAiseadha, C. (2001). "big 'mama' TDB", *Dejavu-L Yahoo Group*, May 28. Retrieved December 14, 2009 from <http://tech.groups.yahoo.com/group/dejavu-l/message/11929?threaded=1&l=1>

- Okunev, A. (2005). "What TM do you work with: small dedicated or Big Momma?", *Dejavu-L Yahoo Group*, February 25. Retrieved on May 4, 2008 from <http://tech.groups.yahoo.com/group/dejavu-l/message/55726?var=1&l=1>
- Oliver, A.; Moré, J.; Climent, S. (2007). *Les technologies de la traducción*. Barcelona: Editorial UOC.
- OQLF (Office québécois de la langue française). (2002). "La fiche terminologique". *La FAQ*. Retrieved October 3, 2011 from <http://www.granddictionnaire.com/BTML/FRA/aide/index.html>
- Parker, B. (1998). *Globalization and Business Practice. Managing Across Boundaries*. London/Thousand Oaks/New Delhi: SAGE Publications Ltd.
- Pavel, S.; Nolet, D. (2001). *Handbook of Terminology*. Ottawa: Public Works and Government Services Canada.
- Pym, A. (2004). *The Moving Text. Localization, translation, and distribution*. Amsterdam/Philadelphia: John Benjamins
- Pym, A. (2011). "What technology does to translating", *Translation & Interpreting* 3(1). Retrieved on December 4, 2011 from www.trans-int.org/index.php/transint/article/viewFile/121/81
- Quah, C. K. (2006). *Translation and Technology*. Basingstoke, UK: Palgrave Macmillan.
- Robinson, D. (2003). *Becoming a Translator. An Introduction to the Theory and Practice of Translation*, 2nd edition. London/New York: Routledge.
- Rondeau, G. (1984). *Introduction à la terminologie*. 2^e édition. Boucherville: Gaëtan Morin.
- Sager, J.C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Samuelsson-Brown, G. (2004). *A Practical Guide for Translators*, 4th revised edition. Frankfurt/Tonawanda/North York: Multilingual Matters Ltd.

- Schonlau, M.; Fricker, R. D.; Elliott, M.N. (2002). *Conducting Research Surveys Via e-Mail and the Web*. Santa Monica: RAND Corporation. Retrieved on April 22, 2012 from <http://site.ebrary.com/lib/oculottawa/Doc?id=10425077&ppg=106>
- SDL Trados (2005). "The History of SDL Trados". *Translationzone*. Retrieved September 23, 2009 from <http://www.translationzone.com/en/about-us/25-year-history-of-trados/>
- SDL Trados (2009a). *Terminology: An End-to-End Perspective. Research Paper*. Retrieved September 6, 2009 from http://www.sdl.com/en/globalization-knowledge-centre/research_results/terminology-an-end-to-end-perspective.asp
- SDLTrados (2009b). *SDL MultiTerm Desktop 2009*. Retrieved December 15, 2009 from <http://www.translationzone.com/en/products/sdlmultitermdesktop/>
- Sofer, M. (2006). *The Translator's Handbook*, 5th edition revised. Rockville, USA: Schreiber Publishing.
- Somers, H. (2003). "Translation Memory Systems." In H. Somers (Ed.), *Computers and Translation: A translator's guide*. Amsterdam/Philadelphia: John Benjamins, pp. 31-47.
- Somers, H. (Ed.) (2003). *Computers and Translation: A translator's guide*, Amsterdam/Philadelphia: John Benjamins.
- STAR (2009). *Transit^{NXT} Context-sensitive translation and localisation*. Retrieved December 15, 2009 from http://star-group.net/downloads/DEU/br_transit-nxt_eng.pdf

Statistics Canada. (2006). *Census of Population*. Statistics Canada catalogue no. 97-559-XCB2006011. Retrieved August 14, 2009 from <http://www12.statcan.gc.ca/english/census06/data/topics/RetrieveProductTable.cfm?TPL=RETR&ALEVEL=3&APATH=3&CATNO=97-559-XCB2006011&DETAIL=0&DIM=&DS=99&FL=0&FREE=0&GAL=0&GC=99&GK=NA&GRP=1&IPS=97-559-CB2006011&METH=0&ORDER=1&PID=92104&PTYPE=88971,97154&RL=0&S=1&ShowAll=No&StartRow=1&SUB=743&Temporal=2006&Theme=74&VID=0&VNAMEE=&VNAMEF=>

Statistics Canada. (2008). *Graduation Rates*. Retrieved November 12, 2011 from <http://www.statcan.gc.ca/pub/81-004-x/def/4068726-eng.htm>

Temmerman, R. (1997). "Questioning the Univocity Ideal. The difference between sociocognitive Terminology and traditional Terminology", *Hermes* 18, pp. 51-90.

Temmerman, R. (2000). *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam/Philadelphia: John Benjamins.

Topping, S. (2000). "Sharing Translation Database Information: Considerations for developing an ethical and viable exchange of data." *Multilingual Computing and Technology* 11(5), pp. 59-61.

Translation Bureau. (2008a). *The Pavel Terminology Tutorial*. Retrieved September 6, 2008 from http://www.termiumplus.gc.ca/didacticiel_tutorial/english/lesson1/index_e.html

Translation Bureau. (2008b). "2.5.2 Recognition of Terminological Units", *The Pavel Terminology Tutorial*. Retrieved December 16, 2009 from http://termiumplus.gc.ca/didacticiel_tutorial/english/lesson2/page2_5_2_e.html

- Translation Bureau. (2008c). "3.4.3 Creating Multilingual Records", *The Pavel Terminology Tutorial*. Retrieved December 18, 2009 from http://www.btb.termiumplus.gc.ca/didacticiel_tutorial/english/lesson3/page3_4_3_e.html
- Trochim. W.M.K. (2006). "Nonprobability Sampling", *Research Methods Knowledge Base*. Retrieved on April 22, 2012 from <http://www.socialresearchmethods.net/kb/sampnon.php>
- Trujillo, A. (1999). *Translation Engines: Techniques for Machine Translation*. London: Springer.
- Warburton, K. (2008). "Terminology Management and Business Opportunities", Power Point presentation, *MultiCorpora's Terminology Management Workshop*, October 8.
- Webb, L.E. (1998). *Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis*, MA Thesis, Monterey Institute of International Studies. Retrieved December 1, 2007 from <http://www.webbsnet.com/translation/thesis.html>
- Wheatley, A. (2003). "A Major Breakthrough for Translator Training," *eContent Localization Resources for Translator Training, Institute of Translation and Interpreting*. Retrieved September 6, 2009 from <http://iti.org.uk/uploadedFiles/surveys/eCoLoRe%20results.pdf>
- Wilss, W. (1999). *Translation and Interpreting in the 20th century. Focus on German*. Amsterdam/Philadelphia: John Benjamins.
- Wittner, J. (2007). "Unexpected ROI from terminology". *Multilingual*, April/May 2007, pp. 1 – 54.
- Wright, S.E.; Budin, G. (1997). *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*. Amsterdam/Philadelphia: John Benjamins.

- Wright, S.E.; Budin, G. (2001). *Handbook of Terminology Management. Volume 2: Application-Oriented Terminology Management*. Amsterdam/Philadelphia: John Benjamins.
- Wright, S.E. (1997). "Term Selection: The Initial Phase of Terminology Management," In S.E. Wright and G. Budin (Eds.), *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*. Amsterdam/Philadelphia: John Benjamins, pp. 13-24.
- Wright, S.E.; Melby, A.K.; Rasmussen, N.; Warburton, K. (2010). "TBX Glossary: A Crosswalk between Termbase and Lexbase Formats," *2010 Association for Machine Translation in the Americas (AMTA) Conference*, Denver, October 31. Retrieved November 6, 2011 from http://amta2010.amtaweb.org/AMTA/papers/TBX-Glossary_2010-10-29.pdf
- Zetsche, J. (2006). "Translation tools come full circle", *Multilingual* 17(1). Retrieved August 21, 2008 from <http://www.multilingual.com/articleDetail.php?id=1175>
- Zetsche, J. (2011). "Advertising in the Tool Kit", *International Writers*. Retrieved on December 5, 2011 from <http://www.internationalwriters.com/toolkit/advertising.html>
- Ye, J. (2011). "Overcoming Challenges to Conducting Online Surveys". In A. Rodney, R. Woods and J.D. Baker (Eds.), *Handbook of Research on Electronic Surveys and Measurements*, Hershey/London: Idea Group Reference, pp.83-89.

APPENDIX A: University of Ottawa Research Ethics Board

Approval Notices



Université d'Ottawa University of Ottawa

Service de subventions de recherche et déontologie Research Grants and Ethics Services

Ethics Approval Notice

Principal Investigator(s) / Supervisor:	Lynne Bowker <input type="text"/>
Co-Investigator(s):	Marta Gomez Plou <input type="text"/>
File Number:	#11-08-15
Title of Research Project:	Investigating the Use of Terminology Management Systems within Translation Environment Tools
Type of Project:	Doctorate thesis
Department and Institution:	School of Translation and Interpretation, Faculty of Arts
Research Ethics Board:	<u>Social Sciences and Humanities</u> Chair: Dr. Peter Beyer
Ethics Approval Date:	January 26, 2009
Expiry Date:	January 25, 2010
Documents Reviewed and Approved:	Protocol
Approval Granted:	Ia (Approval)
Special Conditions:	

This is to confirm that the University of Ottawa Research Ethics Board identified above, which operates in accordance with the Tri-Council Policy Statement and other applicable laws and regulations in Ontario, has examined and approved the application for ethical approval for the above named research project as of the Ethics Approval Date indicated above and subject to the conditions listed the section above entitled "Special Conditions".

During the course of the study the protocol may not be modified without prior written approval from the REB except when necessary to remove subjects from immediate endangerment or when the modification(s) pertain to only administrative or logistical components of the study (e.g. change of telephone number). Investigators must also promptly alert the REB of any changes which increase the risk to participant(s), any changes which considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project and safety of the participant(s). Modifications to the project, information/consent documentation, and/or recruitment documentation, should be submitted to this office for approval using the "Modification to research project" form available at: http://www.rges.uottawa.ca/ethics/application_dwn.asp

Please submit an annual status report to the Protocol Officer 4 weeks before the above-referenced expiry date to either close the file or request a renewal of ethics approval. This document can be found at: http://www.rges.uottawa.ca/ethics/application_dwn.asp



Université d'Ottawa University of Ottawa

Service de subventions de recherche et déontologie Research Grants and Ethics Services

If you have any questions, please do not hesitate to contact the Ethics Office at or by e-mail at

Leslie-Anne Barber
Responsable de la déontologie en recherche
Pour le Président du CÉR en Sciences Sociales et Humanités
Peter Beyer



Université d'Ottawa

University of Ottawa

Bureau d'éthique et d'intégrité de la recherche

Office of Research Ethics and Integrity

Ethics Approval Notice Social Science and Humanities REB

Principal Investigator / Supervisor / Co-investigator(s) / Student(s)

<u>First Name</u>	<u>Last Name</u>	<u>Affiliation</u>	<u>Role</u>
Lynne	Bowker	Arts / Others	Supervisor
Elizabeth	Marshman	Arts / Others	Co-Supervisor
Marta	Gomez Palou	Arts / Translation	Student Researcher

File Number: 04-11-27

Type of Project: PhD Thesis

Title: Integrated Termbases Optimization Survey

Approval Date (mm/dd/yyyy)	Expiry Date (mm/dd/yyyy)	Approval Type
06/21/2011	06/20/2012	Ia

(Ia: Approval, Ib: Approval for initial stage only)

Special Conditions / Comments:

N/A





Université d'Ottawa

Bureau d'éthique et d'intégrité de la recherche

University of Ottawa

Office of Research Ethics and Integrity

This is to confirm that the University of Ottawa Research Ethics Board identified above, which operates in accordance with the Tri-Council Policy Statement and other applicable laws and regulations in Ontario, has examined and approved the application for ethical approval for the above named research project as of the Ethics Approval Date indicated for the period above and subject to the conditions listed the section above entitled "Special Conditions / Comments".

During the course of the study the protocol may not be modified without prior written approval from the REB except when necessary to remove subjects from immediate endangerment or when the modification(s) pertain to only administrative or logistical components of the study (e.g. change of telephone number). Investigators must also promptly alert the REB of any changes which increase the risk to participant(s), any changes which considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project and safety of the participant(s). Modifications to the project, information/consent documentation, and/or recruitment documentation, should be submitted to this office for approval using the "Modification to research project" form available at: http://www.rges.uottawa.ca/ethics/application_dwn.asp

Please submit an annual status report to the Protocol Officer four weeks before the above-referenced expiry date to either close the file or request a renewal of ethics approval. This document can be found at: http://www.rges.uottawa.ca/ethics/application_dwn.asp

If you have any questions, please do not hesitate to contact the Ethics Office at or by e-mail at:

Signature:

Kim Thompson

Protocol Officer for Ethics in Research
For Barbara Graves, Chair of the Social Sciences and Humanities REB

**APPENDIX B: Description of Existing Surveys on
TM Systems Usage and Terminology Management**

LISA 2002 Translation Memory Survey: Translation Memory and Translation Memory Standards by Arle Lommel

This survey carried out by the Localization Industry Standards Association (LISA) aimed to provide a better understanding of the industry's usage of TM tools, perception and understanding of standards, namely Translation Memory eXchange (TMX) and willingness to share TM assets across businesses (Lommel, 2002, p. 3). The survey was distributed online, through the association's website, and obtained 134 responses (Ibid.). Although the survey was open to managers of localization services providers and localization professionals equally, nearly 60% of responses came from managers and executives and only 26% from industry professionals (Ibid.). Educators, engineers and consultants also participated in the survey but in much smaller numbers (Ibid.).

Translation Memory Survey 2003 by Mary Höcker

This survey was designed and distributed by the Institute of Translation & Interpreting (ITI in the United Kingdom) and the *Bundesverband der Dolmetscher und Übersetzer* (BDÜ in Germany, translated into English as the Federal Association of Interpreters and Translators) to their members and other industry contacts such as educational institutions and corporate businesses (Höcker, 2003, p. 1). The survey is part of an eCoLoRe project, funded by the European Union through the Leonardo da Vinci II program. eCoLoRe is a consortium of European translator associations, localization tool developers and training institutions with the goal of developing resources for eContent localization training (Ibid.). The objectives of the survey were to assess TM usage, determine the most frequent subject areas and languages, uncover training requirements, and identify reasons for and against using this type of tool (Ibid.). The survey garnered 208 responses (Ibid.), the majority of which came from

freelancers, with only 8% of respondents being salaried translators (Ibid., p. 2).

2003 *A Major Breakthrough for Translator Training by Alan Wheatley*

This is another report on the eCoLoRe *Translation Memory Survey 2003*. In this case, the results are presented and analyzed by Alan Wheatley, General Secretary of the ITI.

2004 *Portrait of Terminology in Canada by Guy Champagne*

This is an unpublished study that was submitted to the Translation Bureau of Canada in March 2004. The objectives of this study were to assess the presence and impact of in-house and outsourced terminology work in Canadian businesses, to obtain a profile of the terminology sector, to position the role of terminology services as compared to translation or linguistic services, and to identify available terminology resources (Champagne, 2004a, p. 13). This survey targeted managers of departments responsible for terminology services within small and medium sized enterprises (SMEs, i.e. businesses with less than 250 employees) and large corporations (Ibid, p. 15). Over the period of this study, 1724 initial telephone interviews were carried out with managers responsible for terminology services within SMEs and 1431 initial interviews took place with managers responsible for terminology services within large corporations (Ibid, p. 15). Subsequently more detailed interviews were carried out with 133 managers in SMEs and 316 managers in large corporations (Ibid, p. 15). The selection of respondents was carried out randomly by region, company size and sector (Ibid, p. 15).

2004 *The Economic Value of Terminology: An Exploratory Study by Guy Champagne*

This is an unpublished study that was submitted to the Translation Bureau of Canada in

April 2004. This study was complementary to the study presented above. In this case the objectives were to establish the economic value of terminology services within Canadian businesses in terms of revenue, return on investment and cost reduction, as well as to develop performance measurements and grids in order to be able to reproduce this type of study in the future (Champagne, 2004b, p. 13). Two research methods were used: case studies and focus groups. Firstly, 12 case studies were carried out, consisting of an initial telephone interview to establish the eligibility of the terminology service within a company, a questionnaire sent by email, and finally a telephone follow-up to the questionnaire (Champagne, 2004b, p. 15). Companies that participated in the case studies worked in the financial (4), pharmaceutical (2), retail distribution (1), public institution (1) and language services (3) fields, operated across Canada and were based in Québec, Ontario and the prairies provinces (Champagne, 2004b, p. 33). The results obtained in these case studies were validated through two focus group sessions that took place in Ottawa and Montreal with 8 and 12 participants, respectively. As with the previous study, the target audience consisted of managers of departments responsible for terminology services within a company.

LISA 2004 Translation Memory Survey: Translation Memory and Translation Memory Standards by Arle Lommel

This survey was a follow-up to LISA's 2002 *Translation Memory Survey*. The objectives of this survey were to assess current usage of TM tools as well as respondents' plans to use TM tools and TM standards in the future (Lommel, 2004, p. 2). The survey was available online on the Localization Industry Standards Association (LISA) website and it obtained 274 responses, mainly from localization service providers and consumers, with a small proportion coming from tool developers and academic researchers (Ibid). The companies

that responded to this survey had volumes of translation ranging from under one million to over 500 million words per year (Ibid).

LISA 2005 Terminology Management Survey: Terminology Management Practices and Trends by Arle Lommel

This report analyzes the 2004 Terminology Management Survey conducted by LISA's Terminology Special Interest Group. The survey aimed to develop a better understanding of terminology management within the localization industry, such as whether or not terminology was being managed, reasons for not managing terminology, the types of terminology management being carried out, whether terminology management tools were used and what information was collected for each term (Lommel, 2005, p. 1). The survey targeted companies within the localization industry and garnered 81 responses. The majority of respondents were localization service providers, while the second-largest group were users of either localization services or tools (Lommel, 2005, p. 2). Finally, approximately one third of respondents were either localization tool vendors, software companies, educators, consultants, manufacturers or telecommunications companies (Ibid.).

2005 ATIO Survey of Independent Translators by Nancy McInnis and Maha Takla

Nancy McInnis and Maha Takla reported on a survey addressed to all independent (i.e. freelance) translator members of the Association of Translators and Interpreters of Ontario (ATIO). The survey aimed to develop a profile of the freelance translator members of ATIO, including descriptions of age, experience, education, language combination, certification, rates, resources used and personal perception of their professional situation. The survey garnered 193 responses from the 860 ATIO members to whom the survey was distributed

via email invitation.

2005 Translation and Technology: A Study of UK Freelance Translators by Heather Fulford and Joaquin Granell-Zafra

This article reports on the first stage of a project that seeks to investigate the issues surrounding the adoption of various software tools by freelance translators in the United Kingdom (Fulford and Granell-Zafra, 2005, p. 6). The survey enquired about freelance translators' use of different types of software for each activity involved in their workflow, their strategies when adopting software and their general attitude towards software tools (Fulford and Granell-Zafra, 2005, p. 7). The authors distributed the survey by traditional mail and received 391 responses from freelance translators.

2006 OTTIAQ Survey on Rates and Salaries by François Gauthier

François Gauthier reported on the survey carried out by the *Ordre des traducteurs, terminologues et interprètes agréés du Québec* (OTTIAQ) amongst its certified members, candidates for certification and student members (Gauthier, 2006, p. 1). The goal of the survey was to obtain a snapshot of the average rates by service and salaries among members of the association. A total of 493 responses were collected (Ibid.), 60% of which came from freelance translators and 28% from salaried translators (Ibid., p. 2).

2006 Translation Memory Survey. Translation Memory Systems: Enlightening the Users' Perspective by Elina Lagoudaki

This TM survey, unlike many of those described above, did not specifically target corporate users of TM tools but instead reached a sample composed mainly of freelance translators

(Lagoudaki, 2006, p. 7). The goals of this survey were to establish translators' needs, identify tasks related to the use of TM tools, establish profiles of different TM tool users, gain a better knowledge of the different work environments of TM tool users, assess the market penetration of TM tools, uncover the reasons for limited TM use, establish satisfaction levels for different tools, and suggest new ideas for future TM systems (Ibid, p. 8). This survey was made available online via a survey design website (www.surveymonkey.com) and was distributed world-wide through user fora, individual translation and localization companies, associations and organizations, and public institutions that employ translators (Ibid., p. 7). The 874 respondents of the survey came from 54 different countries (Ibid., p. 8). Of these respondents, 90% were translators, and 73% of these worked as freelancers.

2006 eCoLoTrain Results. Translator Training Survey by eCoLoTrain

This survey was carried out by the eCoLoTrain consortium presented above. The survey targeted translator trainers at universities or private companies (eCoLoTrain, p. 4). The goals of the survey were to better understand translator trainers' level of preparation, familiarity and comfort with different computer tools, as well as the approaches they adopted and the challenges they faced when teaching the use of such tools. A total of 86 trainers answered the survey (Ibid.); the responses came from countries across Europe with a heavy proportion coming from Germany and the United Kingdom (Ibid, p. 5). The vast majority of respondents worked for state or private universities while only approximately 7% worked for private companies (Ibid., p. 7).

2006 Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption
by Sarah Dillon and Janet Fraser

This article reports on a research project carried out by Dillon and Fraser which sought to investigate the perception of TM tools by professional translators in the United Kingdom (Dillon and Fraser, 2006, p. 67). The authors conducted an online and email survey among professional translators, students and recent graduates (Ibid., p. 71). The survey was developed and distributed to test three hypotheses: a) that novice translators are more open to using TM tools, b) that TM users perceive these tools more positively than do non-users and c) perceived general proficiency of computer skills is not directly linked to users' perception of TM tools (Ibid., p. 69). The survey garnered 59 responses, 85% of which came from freelance translators and 11% from in-house translators (Ibid., p. 72).

2007 Translation Memory Survey by Institute of Translation & Interpreting (ITI in UK)

The United Kingdom's Institute of Translation & Interpreting conducted this survey, which was open to members and non-members and which addressed translators regardless of whether or not they were users TM tools (ITI, p. 3). The survey was available online, via email and in hard copy (Ibid). Its objectives were to establish current usage of TM tools amongst translators and to identify and qualify training needs with respect to such tools (Ibid.). A total of 163 UK-based translators answered the survey (Ibid.), of whom nearly 84% were freelancers and 11% salaried translators (Ibid., p. 4).

2007 Survey of Salaried Translators by the Association of Translators and Interpreters of Ontario (ATIO)

This survey was carried out by the Association of Translators and Interpreters of Ontario (ATIO) among its members who identified themselves as salaried translators. The objective of the survey was to create a profile of the average salaried translator member of ATIO, describing factors such as gender, age, certification, experience, education, characteristics of

the workplace, industry, language combination and salary (ATIO, 2007). The survey was carried out by means of an email which invited members to complete an online survey. Of the 443 salaried members of ATIO, 119 responded to the survey.

2008 On the Lighter Side: Terminology Results by Nancy McInnis

Nancy McInnis reports on a very small survey carried out amongst ATIO's members regarding their approach to terminology management, focusing on the use of terminology services, creation of personal termbases, client trends regarding the provision of lexicons and members' preferred resources for terminological research (McInnis, 2008). Unfortunately the survey had a very low response rate of only 9% of ATIO members, and the author did not provide any further details about the profile of respondents.

2008 OTTIAQ Survey on Rates and Salaries by François Gauthier

In a follow-on from the 2006 survey, François Gauthier once more reported on a survey carried out by the *Ordre des traducteurs, terminologues et interprètes agréés du Québec* (OTTIAQ) amongst its certified members, candidates for certification and student members (Gauthier, 2008, p. 1). The goal of the survey was to obtain an up-to-date snapshot of the average salaries and rates by service among translators within the association. A total of 532 responses were collected (Ibid.), 62% of which came from freelance translators and 24% from salaried translators (Ibid., p. 2).

2009 The Case for Terminology Management by Nataly Kelly and Donald A. DePalma

Nataly Kelly and Donald A. DePalma conducted a study on behalf of the market research group *Common Sense Advisory*, focusing on the relevance of terminology management. The

goals of the study were to uncover the reasons that motivated companies to undertake this practice, to identify the resulting benefits, and to pinpoint the strategies they followed to implement this new process (Kelly and DePalma, 2009, p. 1). The study involved interviews with terminology managers at 24 European and North American organizations of different sizes and with different terminology management histories and practices (Ibid.).

2009 Terminology: An End-to-End Perspective by SDL

SDL conducted two surveys on the value of terminology management, one of which targeted businesses and the other translators (SDL, 2009, p. 1). The survey of businesses aimed to determine the presence of terminology management and its effects on companies' processes, while the survey of translators investigated how translators perceived the role of terminology management within translation (Ibid.). The 140 respondents to the business survey represented large corporations and in 63% of the cases acted as managers. Of the 194 respondents to the translators' survey, 82% were freelance translators and smaller percentages were in-house translators, language service providers, project managers, terminologists and academics (Ibid.).

2010 Specialised lexicographical resources: a survey of translators' needs by Isabel Durán Muñoz

Isabel Durán Muñoz carried out this survey in order to establish which types of lexicographical resources translators use on a daily basis and to find out which characteristics translators look for in a lexicographical resource in order for it to be considered adequate for meeting their needs. . Duran Muñoz made this survey available in English, Spanish, Italian and German and distributed it online via mailing lists and translation associations. She obtained 402 responses, 62.5% of which came from translators. Durán Muñoz had a special

interest in identifying the particular needs of translators with regard to lexicographical resources, and the needs of professional translators in particular. This is because in the past this type of research had often involved trainee translators.

**APPENDIX C: Use of Terminology Management Systems Integrated
with Translation Environment Tools Survey Questionnaire**

Use of Terminology Management Systems Integrated to Translation

1. Introduction

Welcome to the 2008 survey on the use of terminology management systems integrated with translation environment tools. These tools have been widely commercially available for just over a decade and their use is becoming more and more common across the translation community. The aim of this study is to learn about the community's perception of terminology management systems integrated with translation environment tools as well as to find out more about the approaches taken regarding their use.

Before moving ahead, let's clarify some terms:

- **TRANSLATION ENVIRONMENT TOOLS (TEntTs)** are a type of translation software that integrates in a single collection, or tool suite, a number of computer-aided translation tools intended to facilitate a translator's work. Tools in such a collection could include a translation memory system, a term extraction system and a terminology management system that can interact with each other in the collection.
- **TERMINOLOGY MANAGEMENT SYSTEMS (TMSs)** are software tools that allow its users to store and retrieve terminological information.

If you use a TEntT, your opinion and experience are important for this survey, regardless of whether you actively use the TMS component or whether you even know where to find that particular component of your TEntT.

Participation in this survey is purely voluntary and anonymous. No information that could identify an individual participant is gathered, and IP addresses are not tracked. If at any time you feel that answering a question would compromise your anonymity, you may simply skip the question.

Mandatory questions are identified with an asterisk (*).

*** 1. The data collected will be used to carry out academic research and is likely to be used in the form of pooled data and/or short anonymous excerpt quotations in future publications.**

Do you accept the conditions of this survey?

- Yes, I accept the conditions and want to complete the survey.
- No, I do not want to participate in this survey.

2. Knowledge of English

This survey is only distributed in English. Therefore, in order to participate English will have to be one of your working languages.

***2. Do you have a working knowledge of English?**

- Yes
- No

3. TEnT Tools

***3. Do you use a TEnT such as SDL TRADOS, Déjà Vu, MultiTrans, LogiTerm, Wordfast, Omega-T or similar tool?**

Yes

No

Use of Terminology Management Systems Integrated to Translation

4. Professional Background Information

4. Country of residence:

5. Select the profession that best describes your job position:

- Administrative Assistant
- Company/Section Manager
- Project Manager
- Reviser / Editor
- Other (please specify)
- Technical Writer / Author
- Terminologist
- Translator

6. What language combinations do you work from and into? Please list them in decreasing order of work volume. If you only work in one language, select the applicable language in the From column and N/A in the Into column.

	From	Into
Combination 1:	<input type="text"/>	<input type="text"/>
Combination 2:	<input type="text"/>	<input type="text"/>
Combination 3:	<input type="text"/>	<input type="text"/>
Combination 4:	<input type="text"/>	<input type="text"/>
Combination 5:	<input type="text"/>	<input type="text"/>
Combination 6:	<input type="text"/>	<input type="text"/>

Comments:

7. Which (if any) subject field(s) do you specialize in?

- No specialization
- Education
- Engineering
- Environment
- Other(s) (please specify)
- Finance
- Health
- Information Technologies
- Law
- Marketing
- Pharmaceuticals
- Politics
- Administration

Use of Terminology Management Systems Integrated to Translation

8. Select the work setting that best describes you:

- in-house team of 1 In-house team of 10-49 members Service Provider
 In-house team of 2-9 members In-house team of 50+ members Freelancer

9. Does your team use external contractors?

- Regularly Often Rarely Never Not Applicable

10. How much experience do you have in your field of expertise within the language industry?

- Less than 1 year 1 to 2 years 3 to 5 years 6 to 10 years 11 to 25 years More than 25 years

11. What age range do you belong to?

- 18-24 25-34 35-49 50+

5. Technological Background Information

12. Do you use your TEnT to translate, and if so, how do you use it?

- I don't use it for translation.
- I carry out manual searches.
- I translate interactively (i.e. the TEnT proposes several matches and I choose and adapt the best).
- I pretranslate my texts (i.e. my TEnT automatically replaces the matches it finds and then I edit the resulting text).

13. Which of the TEnTs listed below do you use? (Please choose all that apply)

- | | | |
|---|-------------------------------------|---------------------------------------|
| <input type="checkbox"/> Across | <input type="checkbox"/> LogoPort | <input type="checkbox"/> SDL TRADOS |
| <input type="checkbox"/> Déjà Vu | <input type="checkbox"/> MemoQ | <input type="checkbox"/> Similis |
| <input type="checkbox"/> Fusion | <input type="checkbox"/> MetaTaxis | <input type="checkbox"/> Star Transit |
| <input type="checkbox"/> Heartsome | <input type="checkbox"/> MultiTrans | <input type="checkbox"/> SwordFish |
| <input type="checkbox"/> LogiTerm/LogiTrans | <input type="checkbox"/> Omega-T | <input type="checkbox"/> WordFast |
| <input type="checkbox"/> Other (please specify) | | |

Use of Terminology Management Systems Integrated to Translation

6. Your Main TEnT Background

The rest of the survey will focus on your experience with the TEnT you use more regularly. Please, think of this experience with it and answer the following questions accordingly.

14. Think of how often you use different TEnTs. Which one would you consider your main TEnT?

- | | | |
|--|----------------------------------|------------------------------------|
| <input type="radio"/> Across | <input type="radio"/> LogoPort | <input type="radio"/> SDL TRADOS |
| <input type="radio"/> Déjà Vu | <input type="radio"/> MemoQ | <input type="radio"/> Similis |
| <input type="radio"/> Fusion | <input type="radio"/> MetaTaxis | <input type="radio"/> Star Transit |
| <input type="radio"/> Heartsome | <input type="radio"/> MultiTrans | <input type="radio"/> SwordFish |
| <input type="radio"/> LogiTerm/LogiTrans | <input type="radio"/> Omega-T | <input type="radio"/> WordFast |

15. For how long have you been using your main TEnT?

- Less than 1 year 1 to 2 years 3 to 5 years 6 to 9 years 10 or more

16. Were you able to freely choose your main TEnT?

- Yes, according to my needs (e.g features,budget, etc.)
- No, I adopted my clients' TEnT.
- No, I adopted my employer's TEnT.

17. Have you received formal training on any TEnTs?

- Yes
- No

7. Training

Please answer the questions in this section based on your experience with your main TEnT.

18. What type of formal training on how to use your main TEnT did you receive?

- I took translation technology courses during my studies. I took courses offered by my TEnT provider.
- I took courses offered by industry organisations or professional associations. I took courses offered by my employer.
- Other (please specify)

19. Did any of the formal training that you received cover Terminology Management Systems integrated with TEnTs?

- Yes
- No

8. TMS Training

Please answer the questions in this section based on your experience with your main TEnT.

20. Did the training that you received on the Terminology Management System integrated with your TEnT cover which types of units (most frequent/relevant, nouns, verbs, adjectives, etc.) should be recorded and how?

- Yes
- No

9. Terminology Recording

21. Do you keep any form of term records (e.g. in a notebook, word processor, spreadsheet, terminology management system, etc.)?

- Yes
- No

Use of Terminology Management Systems Integrated to Translation

10. Perception

Please answer the questions in this section based on your experience with your main TEnT.

22. What weight did the Terminology Management System have in the choice of your TEnT?

- Extremely important
- Somewhat important
- Very important
- Not important at all
- Important
- I did not participate in that decision.

23. What is your level of familiarity with and use of the Terminology Management System integrated to your main TEnT?

- I am an expert and have mastered all its features.
- I am comfortable using it, but I have not mastered some advanced features.
- I am comfortable carrying out basic tasks.
- I am uncomfortable, but carry out basic tasks.
- I do not know how to use it.

24. How frequently do you use the Terminology Management System integrated with your TEnT?

- Always
- Very Often
- Often
- Somewhat Often
- Not At All Often
- Never

25. How much time do you spend on an average week looking up word translations, definitions, verifying word spellings, institution names, acronyms or initialisms?

% of time per week

dedicated to terminology

work

Use of Terminology Management Systems Integrated to Translation

26. Below is a list of possible uses of a Terminology Management System. Thinking of your daily use of your Terminology Management System, rank those that apply by order of priority, where 1 is the most important.

	1	2	3	4	5	N/A
To create a glossary or lexicon for a specific field.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To record expressions and their equivalents that required extensive terminological research.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To record expressions and their equivalents that due to their polysemy, variation, connotations or usage can lead to error.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To record expressions and their equivalents that I frequently look up.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To develop a resource that will complement the translation memory database and help the TEnT provide better results when translating a new document.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

11. Planning - Guidelines

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

27. How much planning did you invest in the design and the content of your Terminology Management System records?

- I, or my organization, did not plan. Information is entered as one goes.
- I, or my organization, planned the design and content of the Terminology Management System and have basic general rules for what to record and how.
- I, or my organization, planned the design and content of the Terminology Management System and have very specific guidelines in place on what needs to be recorded and how.

12. Planning - Involvement

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

28. Were you involved in drawing up the guidelines regarding the design and content of your Terminology Management System records?

- Yes
- No

Use of Terminology Management Systems Integrated to Translation

13. Guidelines - Reference Resources

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

29. When planning the design and content of your Terminology Management System records, what resources did you rely on? For those that apply, indicate the degree of importance.

	Very Important	Somewhat Important	Not Too Important	Not At All Important	N/A
Vendor documentation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Industry organisations documentation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Academic works on the topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recommendations by a vendor specialist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specialized courses at academic institutions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specialized courses provided by a professional association or an industry organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other TEnT users' advice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Past experience compiling glossaries in non-TEnT tools	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Past experience compiling glossaries in other TEnT tools	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Existing paper glossaries or dictionaries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please, indicate any other resources you relied on:

14. Guidelines - Approach

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

30. Think of the guidelines you or your organization has established for recording terminology. Do they impose any limitations on the nature of the expressions that can be entered in the database?

- No. I can enter or suggest entering any expression that I consider worthy of being recorded.
- Yes. Expressions must belong to a particular part of speech or a set of parts of speech.
- Yes. Expressions cannot have more than a specific number of words.
- Yes, Other (please specify)

Yes. Expressions must denote a concept.

Yes. Expressions cannot be a synonym of a previously recorded expression.

31. Still thinking of the same guidelines, do they take into account any of the following variables when determining what expressions should be recorded?

- Frequency
- Form variation
- Collocations
- Other (please specify)
- Syntactical agreement with surrounding elements
- None apply

15. Usage - Perspective

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

32. Do you have the rights to create term records in your terminology database(s)?

- Yes, I can create records.
- No, but I can make suggestions which will be evaluated by someone else in my organization.
- No, I can only look-up information.

16. TMS Database Maintenance

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

33. Who creates and feeds your terminology database(s)?

- | | |
|--|--|
| <input type="checkbox"/> Company/Sector Manager(s) | <input type="checkbox"/> Technical Writer(s) / Author(s) |
| <input type="checkbox"/> Subject-field expert(s) | <input type="checkbox"/> Project Manager(s) |
| <input type="checkbox"/> General employee(s) | <input type="checkbox"/> Reviser(s) / Editor(s) |
| <input type="checkbox"/> Sales/Marketing Representative(s) | <input type="checkbox"/> Administrative Assistant(s) |
| <input type="checkbox"/> Terminologist(s) | <input type="checkbox"/> Client(s) |
| <input type="checkbox"/> Translator(s) | <input type="checkbox"/> I don't know |
| <input type="checkbox"/> Other(s) (please specify) | |

17. Usage - Tool

34. Do you keep record of your terminological information with your main TEnT? And, should you keep record of your terminological information with multiple tools, do you use your main TEnT to host your main terminology database or collection?

- Yes.
- No, I use another tool to host my main terminology database or collection.

Use of Terminology Management Systems Integrated to Translation

18. Database Organization

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

35. Do you or your organization store terminology...?

- In one database
- In multiple databases

36. If you use multiple databases, are they divided by ...? (Select all that apply)

- N/A
- Date
- Subject
- Client
- Project
- Language combination
- Other (please specify)

37. If you use only one database, are your records classified by...? (Select all that apply)

- N/A
- Date
- Subject
- Client
- Project
- Language combination
- Other (please specify)

19. Users

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

38. Think of your terminology database(s). Is this database(s) for your own personal use or do you share it/them with other users?

- Some shared, some personal
- Personal
- Shared

20. Shared Terminology Databases

39. Who are the users of the terminology database?

- | | | |
|--|---|---|
| <input type="checkbox"/> Technical Writer(s)/Authors | <input type="checkbox"/> General Public | <input type="checkbox"/> Project Manager(s) |
| <input type="checkbox"/> Translator(s) | <input type="checkbox"/> Reviser(s)/Editor(s) | <input type="checkbox"/> Terminologist(s) |
| <input type="checkbox"/> General Employee(s) | <input type="checkbox"/> Company/Section Manager(s) | <input type="checkbox"/> Client(s) |

Other(s) (please specify)

40. Is there a main user group?

- | | | |
|---|--|---|
| <input type="radio"/> Technical Writer(s)/Authors | <input type="radio"/> Reviser(s)/Editor(s) | <input type="radio"/> Client(s) |
| <input type="radio"/> Translator(s) | <input type="radio"/> Company/Section Manager(s) | <input type="radio"/> There is no main user group |
| <input type="radio"/> General Employee(s) | <input type="radio"/> Project Manager(s) | |
| <input type="radio"/> General Public | <input type="radio"/> Terminologist(s) | |

Use of Terminology Management Systems Integrated to Translation

21. Content Motivation

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

41. Does the ability of your TEnT to automatically look up and insert terminology during the translation process affect your choice of which units to record?

- Yes
- No

42. What weight do the following reasons for storing a unit have in your own decisions to record units?

	Very Important	Somewhat Important	Not Too Important	Not At All Important
It is a key concept of a specialized field.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is an unknown unit that required research.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is a unit whose equivalent you do not know.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is a frequent unit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is a unit that can lead to error (different meanings, connotations, grammatical structure).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is a proper noun (institution, person, document, product, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is a proprietary unit specific to a company/project/subject.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

43. When translating a text, TEnTs draw matches from different resources (translation memory, terminology database, machine translation plug-in, auto-complete function...). When matches are found in multiple resources TEnTs apply certain rules to determine which match takes precedence over the others.

Do you take into account your TEnT's resource prioritization rules when deciding which units to enter in your terminology database?

- Yes
- No
- I am not aware of these rules.

22. Content Sources

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

44. Think of how you find the equivalents for the units you want to record. How much do you rely on your translation memory database(s) to find those equivalents?

- It is / They are the only resource I check.
- It is / They are one of my top resources.
- It is / They are one of the resources I consider.
- It is / They are only a last resort.
- I do not use translation memories as a source for equivalents.

23. Content Type

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

If you are a user with look-up rights only, please answer the questions based on your experience consulting your terminology database records.

45. Think of the units you record in your terminology database. Which if any of the types of units listed below do you store? (Select all that apply)

- | | |
|---|---|
| <input type="checkbox"/> Nouns | <input type="checkbox"/> Telephone and Fax Numbers |
| <input type="checkbox"/> Verbs | <input type="checkbox"/> Proper Nouns (Institutions, People, Places, Products, etc.) |
| <input type="checkbox"/> Adjectives | <input type="checkbox"/> Common Paragraphs |
| <input type="checkbox"/> Websites (URLs) | <input type="checkbox"/> Common Sentences |
| <input type="checkbox"/> Physical Addresses | <input type="checkbox"/> Phrases and other frequent combinations (article + noun, adjective + noun, verb + preposition, etc.) |
| <input type="checkbox"/> E-mail Addresses | |
| <input type="checkbox"/> Other (please specify) | |

Use of Terminology Management Systems Integrated to Translation

46. What information do you include in your term records? Select all that apply and indicate whether they are mandatory or optional.

	Mandatory	Optional	Not Included
Source term	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Target term	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Administrative information (e.g. client, project, date)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Domain(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sub-Domain(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Author of the term record	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grammatical information: part of speech	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grammatical information: gender	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grammatical information: number	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grammatical information: case	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Definition(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Context(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morphological information: inflected forms (gender, number, case, verb tenses, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Syntactical information: structure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lexical information: collocations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Synonym(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Image(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reference material (Web sites, documents, experts)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cross-references (related terms)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Short forms (acronyms, initialisms, symbols, abbreviations)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Source of the term	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Source of the definition	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Source of the context	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Any other sources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Comments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

24. Recording Strategy

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

47. Most Terminology Management Systems allow you to create records with a main entry to which you can assign supporting fields. Which of the units below would you store as main entries or terms?

- | | | |
|--|--|---|
| <input type="checkbox"/> Full forms | <input type="checkbox"/> Preferred spelling of units | <input type="checkbox"/> Websites (URLs) |
| <input type="checkbox"/> Symbols | <input type="checkbox"/> Alternate spellings | <input type="checkbox"/> E-mail Addresses |
| <input type="checkbox"/> Acronyms | <input type="checkbox"/> Base forms of units | <input type="checkbox"/> Physical Addresses |
| <input type="checkbox"/> Initialisms | <input type="checkbox"/> Alternate forms (number, gender, tense) | <input type="checkbox"/> Phraseology (article + noun, adjective + noun, verb + preposition) |
| <input type="checkbox"/> Abbreviations | <input type="checkbox"/> Telephone and Fax numbers | |

48. How do you record synonymic forms? (synonyms, spelling variants, regional variants, symbols, acronyms, initialisms, short forms, etc.)

- Records are organized around a concept and synonymy is only indicated within the concept record in a supporting field.
- Records are organized around a concept and synonyms are entered as terms within the same record.
- Each form is given its own record and synonymy is indicated within each record in a supporting field.
- Each form is given its own record and no synonymic relation is indicated.
- Other (please specify)

49. When recording several synonyms, do you indicate a preferred unit?

- Yes
- No
- N/A

50. What form do you record your units in?

- Always in their base form.
- Whatever form I come across in the text I am translating.
- The form I consider most frequent.
- All forms (number, gender, case, tense).

Use of Terminology Management Systems Integrated to Translation

51. Certain units can take determiners, prepositions or simply tend to appear together with other words in what is known as a collocation. For these units, do you include...? (Select all that apply)

- Combinations/Collocations
- Units on their own

Use of Terminology Management Systems Integrated to Translation

25. Combinations and Collocations

Please, answer the questions in this section based on your experience with your main TEnT or with the TEnT that you use to record terminology.

52. You record determiner, preposition, adverb and adjective + noun or verb combinations. Do you include them in the main entry (term) field or do you add them in a supporting fields such as "collocations", "combinations", "related structures" or "comments"? (Select all that apply)

	As main record entry	In a record field	Not at all
Determiner + noun combinations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Adjective + noun combinations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Verb + preposition combinations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Adverb + verb combinations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you record any other type of combinations (please specify)

53. What determiners or adjectives do you record with a unit?

- Strictly the ones with which the unit appears in my text. Most likely determiners or articles with which the unit can appear.

54. In what forms do you record your combinations of with determiner/adjective + unit?

- In their base form In the form they appear in the text In all gender forms In all number forms In all case forms

55. Do you store multiple forms of combinations in one single record or in separate records?

- Each combination in a separate record All combinations in a single record

Use of Terminology Management Systems Integrated to Translation

56. Do you use any complimentary tool to identify potential collocations or forms of an expression? (Please, select at least one option to be properly directed to the next page.)

- No.
- Yes, paper dictionaries and/or thesauri.
- Yes, an electronic database of words' conceptual-semantic and lexical relations (e.g. WordNet, Visuwords, etc.).
- Yes, the search function in a word processor.
- Yes, the search function in a document management system.
- Yes, my TEnT search function.
- Yes, an Internet search engine.
- Yes, a syntactic analyzer/parser.
- Yes, a concordancer that works on my documents.
- Yes, a concordancer that works on the WWW.
- Other (please specify)

26. Other tool

57. Why do you not use the Terminology Management System integrated with your TEnT?

- I already had developed my terminology database in another system.
- My TEnT Terminology Management System is too complex.
- I never learnt how to use it.
- My TEnT Terminology Management System does not meets my needs.
- Another TMS or tool meets my needs better.
- Other (please specify)

58. What tool do you use to store your terminology information?

- Notebook, index cards or any other paper-format approach
- Stand-alone off-the-shelf Terminology Management System
- Word processor
- In-house Terminology Management System
- Spreadsheet
- Another Terminology Management System integrated with another TEnT
- General database (e.g. Access)
- Other (please specify)

27. No Terminology Recording

59. Why do you not record any terminology?

- I consult existing terminological resources (e.g. paper dictionaries and glossaries, on-line terminology databases and glossaries).
- I consult existing on-line corpora.
- I consult the WWW.
- All terminology information can be found in my repository of past translations (e.g. archive, translation memory, corpora, etc.).
- I do not have the time.
- I have not found a tool that suits my terminological needs.
- I do not bill for my terminological work.
- I do not think it has a value.
- I know all the terminology I need.
- I do not know how.
- I don't know.
- Other (please specify)

28. Recording in the Future

60. Do you think having additional features would encourage you to start recording terminology? If so, what kind?

- No
- Yes (please specify)

61. Do you think receiving (additional) training would encourage you to start recording terminology? If so, what kind?

- No
- Yes (please specify)

62. Do you think having access to more or different kinds of documentation would encourage you to start recording terminology? If so, what kind?

- No
- Yes (please specify)

63. Is there anything else that could encourage you to record terminology?

64. Do you plan to record terminology information in the future?

- Yes
- No
- Don't know

29. Future Tool

65. What tool do you see yourself using to record terminology in the future?

- General database (e.g. Access)
- Word processor
- In-house Terminology Management System
- Terminology Management System integrated with a TEnT
- Spreadsheet
- Stand-alone off-the-shelf Terminology Management System
- Notebook, index cards or any other paper-format approach

30. Other Enhancement Strategies

66. During this survey, we mentioned some strategies to enhance your terminology information retrieval capacity. Do you use any similar strategy we did not mention in the survey? If so, please describe it below.

67. If we have mentioned the strategies you apply, but missed tackling any aspect you consider relevant, please describe these aspects in the text box below.

31. End of Survey

THANKS FOR YOUR INTEREST IN THIS SURVEY!

If you have reached this page after answering negatively to either of questions 1-3, it means that, unfortunately, you do not qualify to participate in this survey. You can click on the **Done** button to exit the survey.

For all other respondents, these are the final questions of the survey. Please, make sure to click on the **Done** button to register your answers and exit the survey.

68. The results of this survey will reveal the actual use the members of the language industry make of Terminology Management Systems and it may also open new doors to explore.

Would you be interested in receiving any subsequent surveys on this topic? If so, please enter your e-mail address in the text box below.

69. If there is any information you want to share and felt the questions did not let you express it clearly, or if you have any additional comments to add, please, feel free to use the text box below.

APPENDIX D: Integrated Termbases Optimization Survey
Questionnaire

Integrated Termbases Optimization Survey

Survey Introduction and Consent

Welcome to the 2011 Integrated Termbases Optimization Survey. This survey is the continuation of a research project started in 2009, which included the Use of Terminology Management Systems Integrated to Translation Environment Tools survey. After collecting information on how users currently perceive, design and use their Terminology Management Systems Integrated to Translation Environment Tools (integrated termbases), we have developed a series of hypotheses on how to optimize the design and use of this type of tool for the purpose of translation. Now we come back to you, the users, to request your input about our hypotheses.

Even if you did not take the 2009 survey, you can still participate in this second stage of research.

CONSENT FORM

Title of the study: Integrated Termbases Optimization Survey

Researchers:

Miss Marta Gomez Palou
School of Translation & Interpretation, Faculty of Arts, University of Ottawa

Lynne Bowker
School of Translation & Interpretation, Faculty of Arts, University of Ottawa

Elizabeth Marshman
School of Translation & Interpretation, Faculty of Arts, University of Ottawa

Invitation to Participate: I am invited to participate in the abovementioned research study conducted by Ms. Gómez Palou, Dr. Bowker and Dr. Marshman.

Purpose of the Study: The purpose of the study is to assess the user acceptance of a series of strategies that will help translators to optimize termbases integrated to translation environment systems (integrated termbases). The end result will be a series of best practices to guide translators on how to best design and build their integrated termbases.

Participation: My participation will consist essentially of completing this survey, which consists of 22 questions. The survey contains a consent question and three sample filtering questions that require a yes or no answer, six demographics questions and 12 content questions. Content questions will present a terminology management strategy to be used in a specific translation scenario and you will be asked to assess the usability of the strategy by answering the related multiple-choice questions. Completing the survey, including reading and accepting this consent form should not take any longer than 30 minutes.

Risks: My participation in this study will entail providing my personal opinion on the terminology management strategies proposed. None of my personal information or my I.P. address will be gathered and participation is completely voluntary.

Benefits: My participation in this study will assist the researchers to validate or discard a series of strategies to optimize integrated termbases. Ultimately, your participation in this research will allow us to provide more reliable guidelines to offer a series of best practices on how to design and build an integrated termbase.

Confidentiality and anonymity: I have received assurance from the researchers that the information I will share will remain strictly confidential. I understand that the contents will be used only for the purpose of assessing the acceptability of the terminology management strategies proposed, that the only people who will have access to the research data are Ms. Gómez Palou, Dr. Bowker and Dr. Marshman, and that results will be published in pooled (aggregate) format. In other words, only overviews of the data, and not individual surveys, will be published.

Anonymity will be protected by offering a completely anonymous survey, during which no personal information will be requested and no I.P. address will be tracked. However, anonymity may be breached if I reveal my identity by providing my name or contact details in one of the open-

Integrated Termbases Optimization Survey

text questions or by contacting the researchers and identifying myself. In any case, no individual quotations will be published and all users will remain anonymous in the results reports.

U.S. Patriot Act: I am aware that this online survey is hosted by "Survey Monkey" which is a web survey company located in the USA. All responses to the survey will be stored and accessed in the USA. This company is subject to U.S. laws, in particular, to the U.S. Patriot Act that allows authorities access to the records of Internet service providers. I understand that my responses to the questions will be stored and accessed in the USA. The security and privacy policy for Survey Monkey can be viewed at <http://www.surveymonkey.com/privacypolicy.aspx>

Conservation of data: The survey results will be kept on a password-protected computer belonging to the researchers at the University of Ottawa and on a password protected Survey Monkey account for a period of 10 years at which time they will be destroyed.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions, without suffering any negative consequences. If I choose to withdraw, all data gathered until the time of withdrawal will be collected and stored as described above. Only the consent and sample filtering questions are mandatory and these will be indicated with an asterisk.

If I have any questions about the study, I may contact the researcher or her supervisors.

If I have any questions regarding the ethical conduct of this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa,

Participants should print a copy of the consent form to keep for their personal records.

*** Do you agree to participate in the above research study conducted by Marta Gómez Palou of the School of Translation and Interpretation – Faculty of Arts (University of Ottawa) which research is under the supervision of Dr. Lynne Bowker and Dr. Elizabeth Marshman?**

Yes, I agree.

No, I do not agree.

Term Definitions

Before we begin, let's clarify some terms:

- **TRANSLATION ENVIRONMENT TOOLS (TEntTs)** are translation software suites that integrate a number of computer-aided translation tools such as a translation memory system, a term extraction system and a terminology management system that interact with each other.
- **TERMINOLOGY MANAGEMENT SYSTEMS (TMSs)** are software tools that allow users to store, organize, and retrieve terminological information.
- **TERMBASES** are the repositories of terminological information created within TMSs.
- **INTERACTIVE TRANSLATION** takes place when the TEntT examines each sentence of a new source text and, sentence by sentence, proposes the matches found in the translation memories and termbases for the translator to assess, adapt and eventually insert as the accepted translation for that sentence.
- **PRETRANSLATION** takes place when the user of a TEntT allows the system to automatically replace any source text sentence or segment with its equivalent found in the translation memory or the termbase.

Sample Filtering Question

***Are you 18-years old or older?**

- Yes
- No

Integrated Termbases Optimization Survey

Sample Filtering Question

***This survey is only available in English. Do you have good reading comprehension in English?**

Yes

No

Integrated Termbases Optimization Survey

Sample Filtering Question

***Do you use a TEnT such as SDL Trados, Déjà Vu, MultiTrans, LogiTerm, Wordfast, MemoQ, Omega-T or a similar tool?**

Yes

No

Integrated Termbases Optimization Survey

General Guidelines

The goal of this survey is to evaluate what type of term record template and term recording strategy would be more useful for a translator working with a TEnT. Throughout the survey you will be presented with a series of term records, source texts and hypothetical TEnT results that have been generated following different design strategies.

The survey exercises require that, unless otherwise specified, you consider each question from the perspective of a translator who uses a generic TEnT (not the specific TEnT you use but a nameless TEnT tool) for personal use with none of the resources being shared.

Some of the samples use English as the source language and French, German and Spanish as target languages. Please note that these languages are used only as example. When answering the questions, do so based on your language combination(s), which may consist of any number of languages.

Again, please, keep in mind that this survey does not enquire about how you use your current TEnT. The aim of the survey is to evaluate which results you would prefer in each of the following scenarios.

LET'S START THE SURVEY!

Integrated Termbases Optimization Survey

Section 1: Recording Synonyms - Interactive Translation

When **translating interactively**, if a unit has multiple source and target equivalents results will be displayed differently depending on how synonyms are recorded.

Option A: If all units are recorded in the same record all available target equivalents are displayed. In general, source language synonyms will not be displayed.

Option B: If units are recorded in separate records by equivalent pair, only the equivalent of the form found in the text will be displayed and none of the other source or target synonyms available for this concept will be displayed.

Consider having to translate this sentence in interactive mode. Your TEnT may propose equivalents for the term **aspirin** as displayed in the Option A or Option B tables below:

*The American Heart Association recommends **aspirin** use for patients who've had a myocardial infarction (heart attack), unstable angina, ischemic stroke (caused by blood clot) or transient ischemic attacks (TIAs or "little strokes"), if not contraindicated.*

American Heart Association. (2011). *Aspirin in Heart Attack and Stroke Prevention*. Retrieved January 30, 2011 from <http://www.americanheart.org/presenter.jhtml?identifier=4456>.

Option A

EN	FR	SP	DE
aspirin	acide acétylsalicylique	ácido acetilsalicílico	Acetylsalicylsäure
	aspirine	aspirina	Aspirin

Option B

EN	FR	SP	D
aspirin	aspirine	aspirina	A

Leaving aside the possibility of other synonyms existing for this particular example, when translating interactively a sentence containing a term such as "aspirin", would you prefer the TEnT to propose the results as displayed in Option A or in Option B?

Option A

Option B

Consider having to translate this sentence in interactive mode. Your TEnT may propose equivalents for the term **GDP** as displayed in the Option A or Option B tables below:

***GDP** at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.*

The World Bank. (2010). *GDP (current US\$)*. Retrieved December 8, 2010 from <http://data.worldbank.org/indicator/NY.GDP.MKTP.CD>.

Integrated Termbases Optimization Survey

Option A

EN	FR	SP	DE
GDP	PIB	PIB	BIP
	produit intérieur brut	producto interior bruto	Bruttoinlandsprodukt

Option B

EN	FR	SP	DE
GDP	PIB	PIB	BIP

When translating interactively a sentence containing an acronym, would you prefer the TEnT to proposed the results as displayed in Option A or in Option B?

Option A

Option B

Integrated Termbases Optimization Survey

Section 2: Recording Synonyms - Pretranslation

When **pretranslating** a sentence containing a unit with multiple target synonyms in the termbase, users may obtain different results depending on how synonyms are recorded.

Option A: Users may choose to record all forms of a concept under a single record. In this case, depending on the system, users will be able to either establish a preferred equivalent to be inserted by default or the first equivalent recorded will be the one inserted.

Option B: Users may choose to record all target language equivalents as a single target term in order for all equivalents to be inserted during pretranslation.

Consider having to pretranslate this sentence. Depending on how synonyms were recorded, your TEnT may insert the target equivalents of the term **global warming** as displayed in the Option A or Option B tables below:

*The main human activities that contribute to **global warming** are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.*

Mastrandrea, M.D.; Schneider, S.H. (2005). *Global warming*. World Book Online Reference Center. 2005. World Book, Inc.

<http://www.worldbookonline.com/wb/Article?id=ar226310>. Retrieved December 8, 2010 from

http://www.nasa.gov/worldbook/global_warming_worldbook.html. " " http://www.nasa.gov/worldbook/global_warming_worldbook.html retrieved December 8, 2010

	Option A	Option B
FR	The main human activities that contribute to réchauffement climatique are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to climatique / réchauffement planétaire planète are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.
SP	The main human activities that contribute to calentamiento de la tierra are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to de la tierra / calentamiento global are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.
DE	The main human activities that contribute to Erdewärmung are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.	The main human activities that contribute to / globale Erwärmung are the burning of fossil fuels (coal, oil, and natural gas) and the clearing of land.

When pretranslating a sentence containing a term with multiple equivalents, would you prefer the TEnT to propose the results as displayed in Option A or in Option B?

Option A

Option B

Integrated Termbases Optimization Survey

Section 3: Recording Non-Term Units

The following questions will focus on whether or not you would consider recording certain types of units useful.

Consider a scenario in which you would have to translate the following text excerpts. Pay particular attention to the unit circled in red and the sample equivalent that would be proposed by the TEnT.

The examples use French as a target language but they can be easily extrapolated to any other target language.

Email addresses that change in the target language

Your support means so much to so many. One of the best ways to help people in need is through a financial donation. You can help people around the world and in your own community. Learn more about current disaster responses that need your support.

Monthly Gift
One-Time Gift

If you have question concerning your tax receipt or your donation please contact



Canadian Red Cross. (2011). *Donate Now!*. Retrieved January 16, 2011 from <http://www.redcross.ca/article.asp?id=43&tid=016>.

Would you consider it useful to create a term record in your TEnT integrated termbase for an email address that changes in the target language?

Yes

No

Integrated Termbases Optimization Survey

Section 3: Recording Non-Term Units

Civic address that changes in the target language

Mail your completed application form, along with the required documents, to:

Citizenship and Immigration Canada
Case Processing Centre – Sydney

EN	FR
Citizenship and Immigration Canada Case Processing Centre – Sydney	Citoyenneté et Immigration Canada Centre de traitement des demandes
<input type="text"/>	<input type="text"/>

Citizenship and Immigration Canada. (2011). *Becoming a citizen: How to apply*. Retrieved January 16, 2011 from <http://www.cic.gc.ca/francais/citoyennete/devenir-demande.asp>.

Would you consider it useful to create a term record in your TEnT integrated termbase for a civic address that changes in the target language?

Yes

No

Integrated Termbases Optimization Survey

Section 3: Recording Non-Term Units

URL that changes in the target language

Bon voyage, but

[...]

BEFORE YOU GO ABROAD, WE ADVISE YOU TO

- Learn the location of Government of Canada offices in the countries you plan to visit and carry the contact information with you (see page 34 or visit www.travel.gc.ca/offices).

EN	FR
www.travel.gc.ca/offices	www.voyage.gc.ca/bureaux

Foreign Affairs and International Trade Canada. (2011). *Travel Report: Estonia*. Retrieved January 20, 2011 from http://www.voyage.gc.ca/countries_pays/report_rapport-eng.asp?id=84000.

Would you consider it useful to create a term record in your TEnT integrated termbase for a URL that changes in the target language?

Yes

No

Section 3: Recording Non-Term Units

Standard text

TRAVEL REPORT

Estonia

[...]

3. SAFETY AND SECURITY

The decision to travel is the sole responsibility of the traveler. The traveler is also responsible for his or her own personal safety. The purpose of this Travel Report is to provide Canadians with up-to-date information to enable them to make well-informed decisions.

Crime

Petty crime, such as pickpocketing, purse-snatching, and mugging occur, particularly in Tallinn's « Old Town » during the summer tourist season. Thieves often work in small groups and target tourists in airports, train stations, and other public areas. Violent crime occurs, but foreigners are not usually targeted. Avoid parks, poorly lit streets, and certain areas of Tallinn, including Kopli, Lasnamäe, and Kadriorg, after dark. Theft of vehicles or their contents is common. Keep vehicles locked and in guarded parking lots overnight.

EN	FR
The decision to travel is the sole responsibility of the traveler. The traveler is also responsible for his or her own personal safety. The purpose of this Travel Report is to provide Canadians with up-to-date information to enable them to make well-informed decisions.	La décision de voyager revient à chaque voyageur. Il incombe également à chacun de veiller à sa sécurité personnelle. Les présents Conseils aux voyageurs ont pour but de fournir des renseignements à jour pour vous aider à prendre des décisions éclairées.

Foreign Affairs and International Trade Canada. (2011). *Bon Voyage, But... Essential Information for Canadian Travellers*. Retrieved January 20, 2011 from http://dsp-psd.pwgsc.gc.ca/collections/collection_2010/maeci-dfait/FR4-5-2010-eng.pdf.

Would you consider it useful to create a term record in your TEnT integrated termbase for a standard text that appears frequently and that must be translated by a standard target translation?

Yes

No

Integrated Termbases Optimization Survey

Section 4: Recording Term Forms

The questions in this section focus on what forms of a term you would consider useful to record.

In this scenario, you are considering adding to your termbase the unit whose different forms appear highlighted in red in the text below.

Please read the text and answer the questions accordingly.

Text containing multiple forms of a term unit

HOW TO MARINATE SAFELY

Marinating or soaking meat, poultry, fish and game in a seasoned liquid, called a marinade, enhances the flavor of the food and often makes it tenderer. However, unless proper food safety techniques are followed, **marinating** can also increase the risk of contamination, which may lead to food borne illness.

[...]

Marinate in the Refrigerator

Always **marinate** foods in the refrigerator, not on the kitchen counter. Many older recipes call for **marinating** at room temperature. But this should never be done as room temperature falls in the Danger Zone between 4°C (40°F) and 60°C (140°F) where bacteria multiply fastest. When a recipe calls for **marinating** at room temperature, increase the **marinating** time in the refrigerator to achieve similar tenderness and taste results.

Do not **marinate** longer than over night. Food has a refrigerated shelf life and **marinating** does not extend that shelf life.

[...]

Marinating Raw Fish

Raw fish is **marinated** in lime juice to make seviche (ceviche), an appetizer popular in Latin America. The action of the acidic lime firms the flesh of the fish and turns it white. But, because the lime juice is diluted by juices from the raw fish, it is not sufficiently acidic to destroy any pathogens which might be present and there is a high risk of food borne illness.

Association of Saskatchewan Home Economists. (2008). *How to marinate safely*. Retrieved February 26, 2011 from http://www.homefamily.net/index.php?/categories/foodnutrition/how_to_marinate_safely/.

When creating the record for this unit, what form or forms would you add to your termbase?

- Only the base form. In this case, as the unit is a verb, the infinitive form: *marinate*
- Multiple forms. In this case, it could be all or any combination or of the following forms: *marinate*, *marinating* or *marinated*.

Section 4: Recording Term Forms

In the last question you indicated that you would record multiple forms of the unit we are considering. Which forms would you record?

- All the forms of the unit that appear in the text. In this example: *marinate*, *marinating* and *marinated*.
- The most frequent form that appears in the text. In this example: *marinating*.
- All the forms of the unit that appear in the text and any other form I can think of. In this example: *marinate*, *marinating*, *marinated* and *marinates*.

Integrated Termbases Optimization Survey

Section 5: TBX-Basic Description

The Localization Industry Standards Association (LISA) has created the TBX-Basic as a simplified version of the international standard TermBase EXchange (TBX), defined in ISO 30042. TBX is an xml coding standard defining how termbases should be structured to facilitate their exchange and including several dozens of default data categories, i.e. fields.

The TBX-Basic defines 3 levels in a term record: concept level, language level and term level. Under each level you may record the fields described in the table below.

Non-mandatory fields can be omitted but **NO additional fields are allowed**.

Mandatory Fields

Term and **language** are mandatory fields, indicated with a red asterisk (*).

Including **at least one** of the three following fields indicated with a double red asterisk (**) is required: **part of speech**, **definition** or **context**

Certain fields can appear under multiple levels.

All fields are text fields except for **term type**, **part of speech**, **gender** and **usage status** which are picklists. The permissible options are listed in the template below. Options can be renamed but **NO additional options are allowed**.

Integrated Termbases Optimization Survey

Here is a sample of a record structure with all fields available on all levels on which they can appear.

Concept Level			
Subject field:			
Image:			
Definition**:			
Source of definition:			
Note:			
Internal cross-reference:	<i>Link to another term in the database</i>		
External cross-reference:	<i>Pointer to explanatory information about the term found in a website, document, etc.</i>		
Creation date:		Created by:	
Last modified date:		Last modified by:	

Language Level			
Language*:			
Definition**:			
Source of definition:			
Note:			
Creation date:		Created by:	
Last modified date:		Last modified by:	

Term Level			
Term*:			
Source of term:			
Term type:	<i>Pick list. Permissible values: fullForm, acronym, abbreviation, shortForm, variant, phrase</i>		
Part of speech**:	<i>Pick list. Permissible values: noun, verb, adjective, adverb, properNoun, other</i>		
Gender:	<i>Pick list. Permissible values: masculine, feminine, neuter, other</i>		
Context**:			
Source of context:			
Usage status:	<i>Pick list. Permissible values: preferred, admitted, notRecommended, obsolete</i>		
Geographical usage:	<i>Countries or locales using the term</i>		
Term location:	<i>Part of a software, publication, packaging where the term frequently occurs</i>		
Internal cross-reference:	<i>Link to another term in the database</i>		
External cross-reference	<i>Pointer to explanatory information about the term found in a website, document, etc.</i>		
Customer:			
Project:			
Note:			
Creation date:		Last modified date:	
Created by:		Last modified by:	

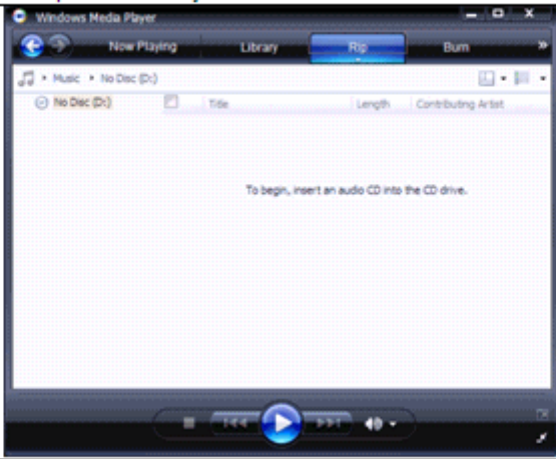
Integrated Termbases Optimization Survey

Section 5: TBX-Basic Record Samples

Applying TBX-Basic you can generate records as simple as this one...

Subject field*:	Computer Memory
Language*:	English
Term*:	rip
Context**:	You can use Windows Media Player 11 to rip songs from your CDs to your computer so that they become files on your computer.

... or as complex as this one.

Subject field*:	Computer Memory		
Image:			
Definition**:	to copy songs from an audio CD to a computer.		
Source of definition:	Microsoft Windows Media Player Help Files: http://windows.microsoft.com/en-US/windows7/Rip-music-from-a-CD		
Note:	It refers to all forms of media.		
Internal cross-reference:	ripping		
External cross-reference:	Microsoft Windows Media Player Help Files: http://windows.microsoft.com/en-US/windows-xp/help/windows-media-player/11/rip-a-cd		
Creation date:	2010/01/05	Created by:	Jane Doe
Last modified date:	2010/01/15	Last modified by:	John Doe

Language*:	English
Note:	Other synonyms may exist but have not yet been recorded.

Term*:	rip
Source of term:	Microsoft Windows Media Player Help Files: http://windows.microsoft.com/en-US/windows-xp/help/windows-media-player/11/rip-a-cd
Term type:	
Part of speech:	verb
Gender:	
Context**:	You can use Windows Media Player 11 to rip songs from your CDs to your computer so that they become files on your computer.
Source of context:	Microsoft Windows Media Player Help Files: http://windows.microsoft.com/en-US/windows-xp/help/windows-media-player/11/rip-a-cd
Usage status:	admitted
Geographical usage:	Canada, United States, UK
Term location:	push button
Customer:	Microsoft

Integrated Termbases Optimization Survey

Customer:	Microsoft		
Project:	Windows Media Player		
Note:	Despite the name, neither the media nor the data in it is damaged after extraction.		
Creation date:	2010/01/05	Created by:	Jane Doe
Last modified date:	2010/01/15	Last modified by:	John Doe

Section 5: Using TBX-Basic for your Term Record Template

Consider the structure (available fields and options, mandatory fields, limitations, etc.) of the TBX-Basic record template.

If you started to use a new TEnT which had the option to use a built-in term record template based on TBX-Basic, where the only modifications allowed would be renaming fields and deleting non-mandatory fields, would you choose to work with this TBX-compliant template?

- Yes, I would base my termbase on the TBX-Basic term record template and would work within its conditions.
- No, I would create my own term record template.

Section 5: Using TBX-Basic for your Term Record Template - Follow Up

You indicated you would use a term record template based on TBX-Basic. Select the reason below that most influenced your decision.

- The record template is based on an industry-approved standard and compliant with an international standard.
- It is easier than creating my own record template from scratch.
- Using a TBX-Basic template ensures that I will be able to easily share my termbase in the future and import other TBX-compliant termbases.
- The TBX-Basic template meets my terminological needs.
- Other (please specify)

Integrated Termbases Optimization Survey

Section 5: Not Using TBX-Basic for your Term Record Template - Follow Up

You indicated you would prefer using your own term record template over a TBX-Basic template. Select the reason below that most influenced your decision.

- TBX-Basic does not include a field that I require in my termbase.
- I do not agree with the term field being mandatory.
- I do not agree with the language field being mandatory.
- I do not agree with either a part of speech, a definition or a context being mandatory.
- TBX-Basic does not include pick list values that I require.
- I am required to use a specific record template structure to maintain compatibility to exchange data with a certain software, client or institution.
- Other

Integrated Termbases Optimization Survey

Demographics

Thanks for hanging in there until this point! The survey is almost over. All that is left is a handful of demographics questions to better manage the sample.

Country of residence:

Profession:

- Translator Terminologist Project Manager
- Other (please specify)

Work setting:

- Freelancer In-house translation service

Main source language:

Main target language:

Have you received any formal education (university courses, certificate, workshop) on terminology theory and/or practices?

- Yes No

AND WE ARE DONE!

Thanks for your interest in this survey. Please make sure to click on the **"Done"** button to register your answers.

**APPENDIX E: Use of Terminology Management Systems Integrated
with Translation Environment Tools Survey Results**

Use of Terminology Management Systems Integrated to Translation Environment Tools



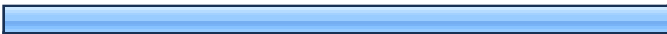
1. The data collected will be used to carry out academic research and is likely to be used in the form of pooled data and/or short anonymous excerpt quotations in future publications. Do you accept the conditions of this survey?

		Response Percent	Response Count
Yes, I accept the conditions and want to complete the survey.		100.0%	104
No, I do not want to participate in this survey.		0.0%	0
		answered question	104
		skipped question	0

2. Do you have a working knowledge of English?




		Response Percent	Response Count
Yes		100.0%	104
No		0.0%	0
		answered question	104
		skipped question	0



3. Do you use a TEnT such as SDL TRADOS, Déjà Vu, MultiTrans, LogiTerm, Wordfast, Omega-T or similar tool?



		Response Percent	Response Count
Yes		100.0%	104
No		0.0%	0
answered question			104
skipped question			0




4. Country of residence:

		Response Percent	Response Count
Afghanistan		0.0%	0
Akrotiri		0.0%	0
Albania		0.0%	0
Algeria		0.0%	0
American Samoa		0.0%	0
Andorra		0.0%	0
Angola		0.0%	0
Anguilla		0.0%	0
Antarctica		0.0%	0
Antigua and Barbuda		0.0%	0
Argentina	▮	1.0%	1
Armenia		0.0%	0
Aruba		0.0%	0
Ashmore and Cartier Islands		0.0%	0
Australia		0.0%	0
Austria	▮	2.1%	2
Azerbaijan		0.0%	0
Bahamas, The		0.0%	0
Bahrain		0.0%	0
Bangladesh		0.0%	0
Barbados		0.0%	0
Bassas da India		0.0%	0
Belarus		0.0%	0

Belgium		2.1%	2
Belize		0.0%	0
Benin		0.0%	0
Bermuda		0.0%	0
Bhutan		0.0%	0
Bolivia		0.0%	0
Bosnia and Herzegovina		0.0%	0
Botswana		0.0%	0
Bouvet Island		0.0%	0
Brazil		3.1%	3
British Indian Ocean Territory		0.0%	0
British Virgin Islands		0.0%	0
Brunei		0.0%	0
Bulgaria		0.0%	0
Burkina Faso		0.0%	0
Burma		0.0%	0
Burundi		0.0%	0
Cambodia		0.0%	0
Cameroon		0.0%	0
Canada		12.4%	12
Cape Verde		0.0%	0
Cayman Islands		0.0%	0
Central African Republic		0.0%	0
Chad		0.0%	0
Chile		0.0%	0



China		0.0%	0
Christmas Island		0.0%	0
Clipperton Island		0.0%	0
Cocos (Keeling) Islands		0.0%	0
Colombia		0.0%	0
Comoros		0.0%	0
Congo, Democratic Republic of the		0.0%	0
Congo, Republic of the		0.0%	0
Cook Islands		0.0%	0
Coral Sea Islands		0.0%	0
Costa Rica		0.0%	0
Cote d'Ivoire		0.0%	0
Croatia		0.0%	0
Cuba		0.0%	0
Cyprus		0.0%	0
Czech Republic		2.1%	2
Denmark		1.0%	1
Dhekelia		0.0%	0
Djibouti		0.0%	0
Dominica		0.0%	0
Dominican Republic		0.0%	0
Ecuador		0.0%	0
Egypt		0.0%	0
El Salvador		0.0%	0
Equatorial Guinea		0.0%	0
Eritrea		0.0%	0






Estonia		0.0%	0
Ethiopia		0.0%	0
Europa Island		0.0%	0
Falkland Islands (Islas Malvinas)		0.0%	0
Faroe Islands		0.0%	0
Fiji		0.0%	0
Finland		0.0%	0
France		13.4%	13
French Guiana		0.0%	0
French Polynesia		0.0%	0
French Southern and Antarctic Lands		0.0%	0
Gabon		0.0%	0
Gambia, The		0.0%	0
Gaza Strip		0.0%	0
Georgia		0.0%	0
Germany		7.2%	7
Ghana		0.0%	0
Gibraltar		0.0%	0
Glorioso Islands		0.0%	0
Greece		2.1%	2
Greenland		0.0%	0
Grenada		0.0%	0
Guadeloupe		0.0%	0
Guam		0.0%	0
Guatemala		0.0%	0




Guernsey		0.0%	0
Guinea		0.0%	0
Guinea-Bissau		0.0%	0
Guyana		0.0%	0
Haiti		0.0%	0
Heard Island and McDonald Islands		0.0%	0
Holy See (Vatican City)		0.0%	0
Honduras		0.0%	0
Hong Kong		0.0%	0
Hungary		0.0%	0
Iceland		0.0%	0
India		0.0%	0
Indonesia		0.0%	0
Iran		0.0%	0
Iraq		0.0%	0
Ireland		1.0%	1
Isle of Man		0.0%	0
Israel		0.0%	0
Italy		2.1%	2
Jamaica		0.0%	0
Jan Mayen		0.0%	0
Japan		2.1%	2
Jersey		0.0%	0
Jordan		0.0%	0
Juan de Nova Island		0.0%	0
Kazakhstan		0.0%	0

Kenya	0.0%	0
Kiribati	0.0%	0
Korea, North	0.0%	0
Korea, South	0.0%	0
Kuwait	0.0%	0
Kyrgyzstan	0.0%	0
Laos	0.0%	0
Latvia	0.0%	0
Lebanon	0.0%	0
Lesotho	0.0%	0
Liberia	0.0%	0
Libya	0.0%	0
Liechtenstein	0.0%	0
Lithuania	0.0%	0
Luxembourg	0.0%	0
Macau	0.0%	0
Macedonia	0.0%	0
Madagascar	0.0%	0
Malawi	0.0%	0
Malaysia	0.0%	0
Maldives	0.0%	0
Mali	0.0%	0
Malta	0.0%	0
Marshall Islands	0.0%	0
Martinique	0.0%	0

Mauritania		0.0%	0
Mauritius		0.0%	0
Mayotte		0.0%	0
Mexico		0.0%	0
Micronesia, Federated States of		0.0%	0
Moldova		0.0%	0
Monaco		0.0%	0
Mongolia		0.0%	0
Montserrat		0.0%	0
Morocco		0.0%	0
Mozambique		0.0%	0
Namibia		0.0%	0
Nauru		0.0%	0
Navassa Island		0.0%	0
Nepal		0.0%	0
Netherlands	🇳🇱	1.0%	1
Netherlands Antilles		0.0%	0
New Caledonia		0.0%	0
New Zealand		0.0%	0
Nicaragua		0.0%	0
Niger		0.0%	0
Nigeria		0.0%	0
Niue		0.0%	0
Norfolk Island		0.0%	0
Northern Mariana Islands		0.0%	0
Norway	🇳🇴	1.0%	1







Oman		0.0%	0
Pakistan		0.0%	0
Palau		0.0%	0
Panama		0.0%	0
Papua New Guinea		0.0%	0
Paracel Islands		0.0%	0
Paraguay		0.0%	0
Peru		0.0%	0
Philippines		0.0%	0
Pitcairn Islands		0.0%	0
Poland		0.0%	0
Portugal		6.2%	6
Puerto Rico		0.0%	0
Qatar		0.0%	0
Reunion		0.0%	0
Romania		3.1%	3
Russia		0.0%	0
Rwanda		0.0%	0
Saint Helena		0.0%	0
Saint Kitts and Nevis		0.0%	0
Saint Lucia		0.0%	0
Saint Pierre and Miquelon		0.0%	0
Saint Vincent and the Grenadines		0.0%	0
Samoa		0.0%	0
San Marino		0.0%	0
Sao Tome and Principe		0.0%	0

Saudi Arabia		0.0%	0
Senegal		0.0%	0
Serbia and Montenegro		0.0%	0
Seychelles		0.0%	0
Sierra Leone		0.0%	0
Singapore		0.0%	0
Slovakia		0.0%	0
Slovenia		1.0%	1
Solomon Islands		0.0%	0
Somalia		0.0%	0
South Africa		1.0%	1
South Georgia and the South Sandwich Islands		0.0%	0
Spain		12.4%	12
Spratly Islands		0.0%	0
Sri Lanka		0.0%	0
Sudan		0.0%	0
Suriname		0.0%	0
Svalbard		0.0%	0
Swaziland		0.0%	0
Sweden		1.0%	1
Switzerland		1.0%	1
Syria		0.0%	0
Taiwan		0.0%	0
Tajikistan		0.0%	0
Tanzania		0.0%	0

Thailand		0.0%	0
Timor-Leste		0.0%	0
Togo		0.0%	0
Tokelau		0.0%	0
Tonga		0.0%	0
Trinidad and Tobago		0.0%	0
Tromelin Island		0.0%	0
Tunisia		0.0%	0
Turkey		0.0%	0
Turkmenistan		0.0%	0
Turks and Caicos Islands		0.0%	0
Tuvalu		0.0%	0
Uganda		0.0%	0
Ukraine		1.0%	1
United Arab Emirates		0.0%	0
United Kingdom		8.2%	8
United States		11.3%	11
Uruguay		0.0%	0
Uzbekistan		0.0%	0
Vanuatu		0.0%	0
Venezuela		0.0%	0
Vietnam		0.0%	0
Virgin Islands		0.0%	0
Wake Island		0.0%	0
Wallis and Futuna		0.0%	0

West Bank	0.0%	0
Western Sahara	0.0%	0
Yemen	0.0%	0
Zambia	0.0%	0
Zimbabwe	0.0%	0
answered question		97
skipped question		7

5. Select the profession that best describes your job position:

		Response Percent	Response Count
Administrative Assistant		0.0%	0
Company/Section Manager		5.8%	6
Translator		74.0%	77
Terminologist		7.7%	8
Technical Writer / Author		0.0%	0
Reviser / Editor		1.9%	2
Project Manager		3.8%	4
Other (please specify)		6.7%	7
answered question			104
skipped question			0

6. What language combinations do you work from and into? Please list them in decreasing order








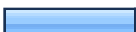





From

	N/A	Amharic	Arabic	Armenian	Basque	Bengali
Combination 1:	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 2:	1.4% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 3:	2.9% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 4:	5.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 5:	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 6:	20.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)







Into

	N/A	Amharic	Arabic	Armenian	Basque	Bengali
Combination 1:	3.9% (4)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 2:	1.4% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 3:	3.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 4:	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 5:	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Combination 6:	20.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)






7. Which (if any) subject field(s) do you specialize in?

		Response Percent	Response Count
No specialization		12.6%	13
Education		5.8%	6
Engineering		35.0%	36
Environment		16.5%	17
Finance		18.4%	19
Health		18.4%	19
Information Technologies		43.7%	45
Law		19.4%	20
Marketing		20.4%	21
Pharmaceuticals		13.6%	14
Politics		4.9%	5
Administration		7.8%	8
Other(s) (please specify)		29.1%	30
		answered question	103
		skipped question	1

8. Select the work setting that best describes you:

		Response Percent	Response Count
in-house team of 1		13.5%	14
In-house team of 2-9 members		21.2%	22
In-house team of 10-49 members		10.6%	11
In-house team of 50+ members		1.9%	2
Service Provider		4.8%	5
Freelancer		48.1%	50
		answered question	104
		skipped question	0

9. Does your team use external contractors?

		Response Percent	Response Count
Regularly		31.1%	32
Often		11.7%	12
Rarely		24.3%	25
Never		7.8%	8
Not Applicable		25.2%	26
		answered question	103
		skipped question	1





10. How much experience do you have in your field of expertise within the language industry?

		Response Percent	Response Count
Less than 1 year		1.0%	1
1 to 2 years		4.8%	5
3 to 5 years		15.4%	16
6 to 10 years		25.0%	26
11 to 25 years		40.4%	42
More than 25 years		13.5%	14
answered question			104
skipped question			0






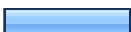










11. What age range do you belong to?

		Response Percent	Response Count
18-24		1.9%	2
25-34		25.2%	26
35-49		42.7%	44
50+		30.1%	31
answered question			103
skipped question			1












12. Do you use your TEnT to translate, and if so, how do you use it?

		Response Percent	Response Count
I don't use it for translation.		6.7%	7
I carry out manual searches.		15.4%	16
I translate interactively (i.e. the TEnT proposes several matches and I choose and adapt the best).		78.8%	82
I pretranslate my texts (i.e. my TEnT automatically replaces the matches it finds and then I edit the resulting text).		29.8%	31
		answered question	104
		skipped question	0




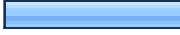

13. Which of the TEnTs listed below do you use? (Please choose all that apply)

		Response Percent	Response Count
Fusion		1.0%	1
SDL TRADOS		68.0%	70
LogiTerm/LogiTrans		8.7%	9
MultiTrans		9.7%	10
Déjà Vu		24.3%	25
WordFast		18.4%	19
Star Transit		17.5%	18
Omega-T		7.8%	8
Similis		1.9%	2
Across		8.7%	9
MetaTaxis		1.9%	2
Heartsome		3.9%	4
LogoPort		3.9%	4
MemoQ		9.7%	10
SwordFish		1.0%	1
Other (please specify)		10.7%	11
answered question			103
skipped question			1




14. Think of how often you use different TEnTs. Which one would you consider your main TEnT?

		Response Percent	Response Count
Fusion		1.0%	1
SDL TRADOS		49.0%	49
LogiTerm/LogiTrans		3.0%	3
MultiTrans		8.0%	8
Déjà Vu		14.0%	14
WordFast		8.0%	8
Star Transit		9.0%	9
Omega-T		3.0%	3
Similis		1.0%	1
Across		1.0%	1
MetaTaxis		0.0%	0
Heartsome		0.0%	0
LogoPort		0.0%	0
MemoQ		3.0%	3
SwordFish		0.0%	0
answered question			100
skipped question			4



15. For how long have you been using your main TEnT?

		Response Percent	Response Count
Less than 1 year		10.7%	11
1 to 2 years		12.6%	13
3 to 5 years		31.1%	32
6 to 9 years		26.2%	27
10 or more		19.4%	20
		answered question	103
		skipped question	1


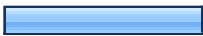



16. Were you able to freely choose your main TEnT?

		Response Percent	Response Count
Yes, according to my needs (e.g features,budget, etc.)		70.9%	73
No, I adopted my clients' TEnT.		11.7%	12
No, I adopted my employer's TEnT.		17.5%	18
		answered question	103
		skipped question	1



17. Have you received formal training on any TEnTs?

		Response Percent	Response Count
Yes		50.5%	52
No		49.5%	51
answered question			103
skipped question			1


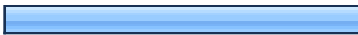
18. What type of formal training on how to use your main TEnT did you receive?

		Response Percent	Response Count
I took translation technology courses during my studies.		27.5%	14
I took courses offered by industry organisations or professional associations.		29.4%	15
I took courses offered by my TEnT provider.		52.9%	27
I took courses offered by my employer.		25.5%	13
Other (please specify)		9.8%	5
answered question			51
skipped question			53



19. Did any of the formal training that you received cover Terminology Management Systems integrated with TEnTs?

		Response Percent	Response Count
Yes		62.7%	32
No		37.3%	19
answered question			51
skipped question			53

20. Did the training that you received on the Terminology Management System integrated with your TEnT cover which types of units (most frequent/relevant,nouns, verbs, adjectives, etc.) should be recorded and how?

		Response Percent	Response Count
Yes		46.7%	14
No		53.3%	16
answered question			30
skipped question			74

21. Do you keep any form of term records (e.g. in a notebook, word processor, spreadsheet, terminology management system, etc.)?

		Response Percent	Response Count
Yes		86.4%	19
No		13.6%	3
answered question			22
skipped question			82







22. What weight did the Terminology Management System have in the choice of your TEnT?

		Response Percent	Response Count
Extremely important		17.2%	17
Very important		20.2%	20
Important		15.2%	15
Somewhat important		17.2%	17
Not important at all		19.2%	19
I did not participate in that decision.		11.1%	11
answered question			99
skipped question			5

23. What is your level of familiarity with and use of the Terminology Management System integrated to your main TEnT?

		Response Percent	Response Count
I am an expert and have mastered all its features.		30.0%	30
I am comfortable using it, but I have not mastered some advanced features.		40.0%	40
I am comfortable carrying out basic tasks.		21.0%	21
I am uncomfortable, but carry out basic tasks.		4.0%	4
I do not know how to use it.		5.0%	5
answered question			100
skipped question			4

24. How frequently do you use the Terminology Management System integrated with your TEnT?

		Response Percent	Response Count
Always		40.0%	40
Very Often		19.0%	19
Often		17.0%	17
Somewhat Often		5.0%	5
Not At All Often		14.0%	14
Never		5.0%	5
answered question			100
skipped question			4




25. How much time do you spend on an average week looking up word translations, definitions, verifying word spellings, institution names, acronyms or initialisms?

	Response Average	Response Total	Response Count
% of time per week dedicated to terminology work	24.39	2,317	95
answered question			95
skipped question			9



26. Below is a list of possible uses of a Terminology Management System. Thinking of your daily use of your Terminology Management System, rank those that apply by order of priority, where 1 is the most important.

	1	2	3	4	5	N/A	Rating Average	Response Count
To create a glossary or lexicon for a specific field.	45.5% (45)	17.2% (17)	11.1% (11)	10.1% (10)	8.1% (8)	8.1% (8)	3.89	99
To record expressions and their equivalents that required extensive terminological research.	47.9% (46)	24.0% (23)	11.5% (11)	7.3% (7)	5.2% (5)	4.2% (4)	4.07	96
To record expressions and their equivalents that due to their polysemy, variation, connotations or usage can lead to error.	17.7% (17)	25.0% (24)	13.5% (13)	12.5% (12)	22.9% (22)	8.3% (8)	3.02	96
To record expressions and their equivalents that I frequently look up.	33.0% (32)	22.7% (22)	24.7% (24)	9.3% (9)	5.2% (5)	5.2% (5)	3.73	97
To develop a resource that will complement the translation memory database and help the TEnT provide better results when translating a new document.	44.9% (44)	21.4% (21)	11.2% (11)	9.2% (9)	7.1% (7)	6.1% (6)	3.93	98
						Other (please specify)		15
						answered question		99
						skipped question		5

27. How much planning did you invest in the design and the content of your Terminology Management System records?

		Response Percent	Response Count
I, or my organization, did not plan. Information is entered as one goes.		53.1%	52
I, or my organization, planned the design and content of the Terminology Management System and have very specific guidelines in place on what needs to be recorded and how.		13.3%	13
I, or my organization, planned the design and content of the Terminology Management System and have basic general rules for what to record and how.		33.7%	33
		answered question	98
		skipped question	6

28. Were you involved in drawing up the guidelines regarding the design and content of your Terminology Management System records?

		Response Percent	Response Count
Yes		81.3%	39
No		18.8%	9
		answered question	48
		skipped question	56







29. When planning the design and content of your Terminology Management System records, what resources did you rely on? For those that apply, indicate the degree of importance.

	Very Important	Somewhat Important	Not Too Important	Not At All Important	N/A	Response Count
Vendor documentation	45.7% (16)	37.1% (13)	14.3% (5)	0.0% (0)	2.9% (1)	35
Industry organisations documentation	31.3% (10)	28.1% (9)	18.8% (6)	12.5% (4)	9.4% (3)	32
Academic works on the topic	24.2% (8)	24.2% (8)	12.1% (4)	24.2% (8)	15.2% (5)	33
Recommendations by a vendor specialist	15.2% (5)	33.3% (11)	18.2% (6)	15.2% (5)	18.2% (6)	33
Specialized courses at academic institutions	23.5% (8)	23.5% (8)	11.8% (4)	17.6% (6)	23.5% (8)	34
Specialized courses provided by a professional association or an industry organization	21.9% (7)	25.0% (8)	21.9% (7)	12.5% (4)	18.8% (6)	32
Other TEnT users' advice	17.6% (6)	50.0% (17)	23.5% (8)	5.9% (2)	2.9% (1)	34
Past experience compiling glossaries in non-TEnT tools	38.2% (13)	35.3% (12)	17.6% (6)	0.0% (0)	8.8% (3)	34
Past experience compiling glossaries in other TEnT tools	32.4% (11)	35.3% (12)	14.7% (5)	2.9% (1)	14.7% (5)	34
Existing paper glossaries or dictionaries	38.2% (13)	32.4% (11)	14.7% (5)	11.8% (4)	2.9% (1)	34







Please, indicate any other resources you relied on: 5

answered question	38
skipped question	66




30. Think of the guidelines you or your organization has established for recording terminology. Do they impose any limitations on the nature of the expressions that can be entered in the database?

		Response Percent	Response Count
No. I can enter or suggest entering any expression that I consider worthy of being recorded.		68.9%	31
Yes. Expressions must belong to a particular part of speech or a set of parts of speech.		4.4%	2
Yes. Expressions cannot have more than a specific number of words.		6.7%	3
Yes. Expressions must denote a concept.		31.1%	14
Yes. Expressions cannot be a synonym of a previously recorded expression.		8.9%	4
Yes, Other (please specify)		2.2%	1
		answered question	45
		skipped question	59



31. Still thinking of the same guidelines, do they take into account any of the following variables when determining what expressions should be recorded?

		Response Percent	Response Count
Frequency		47.7%	21
Form variation		36.4%	16
Collocations		38.6%	17
Syntactical agreement with surrounding elements		25.0%	11
None apply		27.3%	12
Other (please specify)		2.3%	1
answered question			44
skipped question			60

32. Do you have the rights to create term records in your terminology database(s)?

		Response Percent	Response Count
Yes, I can create records.		95.8%	91
No, but I can make suggestions which will be evaluated by someone else in my organization.		3.2%	3
No, I can only look-up information.		1.1%	1
answered question			95
skipped question			9

33. Who creates and feeds your terminology database(s)?

		Response Percent	Response Count
Company/Sector Manager(s)		0.0%	0
Subject-field expert(s)		0.0%	0
General employee(s)		0.0%	0
Sales/Marketing Representative(s)		0.0%	0
Terminologist(s)		0.0%	0
Translator(s)		100.0%	1
Technical Writer(s) / Author(s)		0.0%	0
Project Manager(s)		0.0%	0
Reviser(s) / Editor(s)		0.0%	0
Administrative Assistant(s)		0.0%	0
Client(s)		100.0%	1
I don't know		0.0%	0
Other(s) (please specify)		0.0%	0
answered question			1
skipped question			103

34. Do you keep record of your terminological information with your main TEnT? And, should you keep record of your terminological information with multiple tools, do you use your main TEnT to host your main terminology database or collection?

		Response Percent	Response Count
Yes.		78.1%	75
No, I use another tool to host my main terminology database or collection.		21.9%	21
		answered question	96
		skipped question	8

35. Do you or your organization store terminology...?

		Response Percent	Response Count
In one database		18.4%	14
In multiple databases		81.6%	62
		answered question	76
		skipped question	28

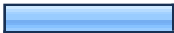


36. If you use multiple databases, are they divided by ...? (Select all that apply)

		Response Percent	Response Count
N/A		15.5%	11
Date		0.0%	0
Subject		63.4%	45
Client		45.1%	32
Project		19.7%	14
Language combination		26.8%	19
Other (please specify)		7.0%	5
answered question			71
skipped question			33










37. If you use only one database, are your records classified by...? (Select all that apply)

		Response Percent	Response Count
N/A		54.5%	24
Date		15.9%	7
Subject		31.8%	14
Client		15.9%	7
Project		20.5%	9
Language combination		25.0%	11
Other (please specify)		0.0%	0
answered question			44
skipped question			60






38. Think of your terminology database(s). Is this database(s) for your own personal use or do you share it/them with other users?

		Response Percent	Response Count
Some shared, some personal		25.0%	19
Personal		40.8%	31
Shared		34.2%	26
answered question			76
skipped question			28

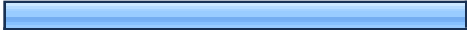

39. Who are the users of the terminology database?

		Response Percent	Response Count
Technical Writer(s)/Authors		13.3%	6
Translator(s)		95.6%	43
General Employee(s)		15.6%	7
General Public		8.9%	4
Reviser(s)/Editor(s)		51.1%	23
Company/Section Manager(s)		6.7%	3
Project Manager(s)		42.2%	19
Terminologist(s)		37.8%	17
Client(s)		22.2%	10
Other(s) (please specify)			2
answered question			45
skipped question			59

40. Is there a main user group?

		Response Percent	Response Count
Technical Writer(s)/Authors		0.0%	0
Translator(s)		74.4%	32
General Employee(s)		0.0%	0
General Public		0.0%	0
Reviser(s)/Editor(s)		2.3%	1
Company/Section Manager(s)		0.0%	0
Project Manager(s)		7.0%	3
Terminologist(s)		4.7%	2
Client(s)		0.0%	0
There is no main user group		11.6%	5
answered question			43
skipped question			61

41. Does the ability of your TEnT to automatically look up and insert terminology during the translation process affect your choice of which units to record?




		Response Percent	Response Count
Yes		69.4%	50
No		30.6%	22
answered question			72
skipped question			32

42. What weight do the following reasons for storing a unit have in your own decisions to record units?





	Very Important	Somewhat Important	Not Too Important	Not At All Important	Response Count
It is a key concept of a specialized field.	81.1% (60)	16.2% (12)	2.7% (2)	0.0% (0)	74
It is an unknown unit that required research.	69.9% (51)	26.0% (19)	2.7% (2)	1.4% (1)	73
It is a unit whose equivalent you do not know.	49.3% (35)	38.0% (27)	8.5% (6)	4.2% (3)	71
It is a frequent unit.	56.9% (41)	22.2% (16)	16.7% (12)	4.2% (3)	72
It is a unit that can lead to error (different meanings, connotations, grammatical structure).	55.6% (40)	26.4% (19)	18.1% (13)	0.0% (0)	72
It is a proper noun (institution, person, document, product, etc.).	32.4% (24)	32.4% (24)	29.7% (22)	5.4% (4)	74
It is a proprietary unit specific to a company/project/subject.	58.1% (43)	28.4% (21)	10.8% (8)	2.7% (2)	74
answered question					74
skipped question					30

43. When translating a text, TEnTs draw matches from different resources (translation memory, terminology database, machine translation plug-in, auto-complete function...). When matches are found in multiple resources TEnTs apply certain rules to determine which match takes precedence over the others.

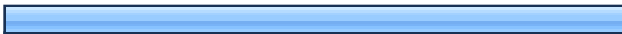









Do you take into account your TEnT's resource prioritization rules when deciding which units to enter in your terminology database?

		Response Percent	Response Count
Yes		44.0%	33
No		22.7%	17
I am not aware of these rules.		33.3%	25
		answered question	75
		skipped question	29

44. Think of how you find the equivalents for the units you want to record. How much do you rely on your translation memory database(s) to find those equivalents?

		Response Percent	Response Count
It is / They are the only resource I check.		0.0%	0
It is / They are one of my top resources.		39.5%	30
It is / They are one of the resources I consider.		46.1%	35
It is / They are only a last resort.		5.3%	4
I do not use translation memories as a source for equivalents.		9.2%	7
		answered question	76
		skipped question	28

45. Think of the units you record in your terminology database. Which if any of the types of units listed below do you store? (Select all that apply)












		Response Percent	Response Count
Nouns		93.2%	68
Verbs		82.2%	60
Adjectives		78.1%	57
Websites (URLs)		8.2%	6
Physical Addresses		0.0%	0
E-mail Addresses		1.4%	1
Telephone and Fax Numbers		0.0%	0
Proper Nouns (Institutions, People, Places, Products, etc.)		63.0%	46
Common Paragraphs		16.4%	12
Common Sentences		32.9%	24
Phrases and other frequent combinations (article + noun, adjective + noun, verb + preposition, etc.)		72.6%	53
Other (please specify)		9.6%	7
answered question			73
skipped question			31

46. What information do you include in your term records? Select all that apply and indicate whether they are mandatory or optional.






	Mandatory	Optional	Not Included	Response Count
Source term	98.7% (75)	1.3% (1)	0.0% (0)	76
Target term	96.1% (73)	3.9% (3)	0.0% (0)	76
Administrative information (e.g. client, project, date)	35.6% (26)	49.3% (36)	15.1% (11)	73
Domain(s)	38.0% (27)	42.3% (30)	19.7% (14)	71
Sub-Domain(s)	10.6% (7)	51.5% (34)	37.9% (25)	66
Author of the term record	38.6% (27)	32.9% (23)	28.6% (20)	70
Grammatical information: part of speech	10.1% (7)	47.8% (33)	42.0% (29)	69
Grammatical information: gender	5.8% (4)	49.3% (34)	44.9% (31)	69
Grammatical information: number	4.4% (3)	44.1% (30)	51.5% (35)	68
Grammatical information: case	3.0% (2)	40.3% (27)	56.7% (38)	67
Definition(s)	15.1% (11)	69.9% (51)	15.1% (11)	73
Context(s)	20.5% (15)	67.1% (49)	12.3% (9)	73
Morphological information: inflected forms (gender, number, case, verb tenses, etc.)	1.5% (1)	41.8% (28)	56.7% (38)	67
Syntactical information: structure	0.0% (0)	37.3% (25)	62.7% (42)	67
Lexical information: collocations	3.0% (2)	53.7% (36)	43.3% (29)	67
Synonym(s)	5.7% (4)	78.6% (55)	15.7% (11)	70
Image(s)	1.4% (1)	45.7% (32)	52.9% (37)	70
Reference material (Web sites, documents, experts)	7.1% (5)	65.7% (46)	27.1% (19)	70
Cross-references (related terms)	4.4% (3)	64.7% (44)	30.9% (21)	68

Short forms (acronyms, initialisms, symbols, abbreviations)	15.7% (11)	68.6% (48)	15.7% (11)	70
Source of the term	32.9% (23)	41.4% (29)	25.7% (18)	70
Source of the definition	24.3% (17)	45.7% (32)	30.0% (21)	70
Source of the context	22.4% (15)	43.3% (29)	34.3% (23)	67
Any other sources	6.3% (4)	50.8% (32)	42.9% (27)	63
Comments	1.5% (1)	81.5% (53)	16.9% (11)	65
			Other (please specify)	6
answered question				76
skipped question				28




47. Most Terminology Management Systems allow you to create records with a main entry to which you can assign supporting fields. Which of the units below would you store as main entries or terms?

		Response Percent	Response Count
Full forms		92.8%	64
Symbols		10.1%	7
Acronyms		47.8%	33
Initialisms		8.7%	6
Abbreviations		49.3%	34
Preferred spelling of units		23.2%	16
Alternate spellings		17.4%	12
Base forms of units		23.2%	16
Alternate forms (number, gender, tense)		13.0%	9
Telephone and Fax numbers		0.0%	0
Websites (URLs)		2.9%	2
E-mail Addresses		0.0%	0
Physical Adresses		0.0%	0
Phraseology (article + noun, adjective + noun, verb + preposition)		36.2%	25
answered question			69
skipped question			35



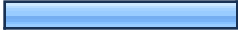

48. How do you record synonymic forms? (synonyms, spelling variants, regional variants, symbols, acronyms, initialisms, short forms, etc.)

		Response Percent	Response Count
Records are organized around a concept and synonymy is only indicated within the concept record in a supporting field.		17.1%	12
Records are organized around a concept and synonyms are entered as terms within the same record.		37.1%	26
Each form is given its own record and synonymy is indicated within each record in a supporting field.		15.7%	11
Each form is given its own record and no synonymic relation is indicated.		25.7%	18
Other (please specify)		4.3%	3
		answered question	70
		skipped question	34


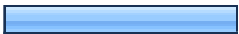
49. When recording several synonyms, do you indicate a preferred unit?

		Response Percent	Response Count
Yes		43.7%	31
No		39.4%	28
N/A		16.9%	12
		answered question	71
		skipped question	33

50. What form do you record your units in?

		Response Percent	Response Count
Always in their base form.		44.0%	33
Whatever form I come across in the text I am translating.		17.3%	13
The form I consider most frequent.		34.7%	26
All forms (number, gender, case, tense).		4.0%	3
		answered question	75
		skipped question	29



51. Certain units can take determiners, prepositions or simply tend to appear together with other words in what is known as a collocation. For these units, do you include...? (Select all that apply)

		Response Percent	Response Count
Combinations/Collocations		65.3%	47
Units on their own		34.7%	25
		answered question	72
		skipped question	32






52. You record determiner, preposition, adverb and adjective + noun or verb combinations. Do you include them in the main entry (term) field or do you add them in a supporting fields such as "collocations", "combinations", "related structures" or "comments"? (Select all that apply)

	As main record entry	In a record field	Not at all	Response Count
Determiner + noun combinations	59.6% (28)	27.7% (13)	12.8% (6)	47
Adjective + noun combinations	62.2% (28)	33.3% (15)	4.4% (2)	45
Verb + preposition combinations	58.1% (25)	32.6% (14)	9.3% (4)	43
Adverb + verb combinations	52.4% (22)	28.6% (12)	19.0% (8)	42
If you record any other type of combinations (please specify)				0
answered question				47
skipped question				57

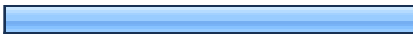

53. What determiners or adjectives do you record with a unit?

		Response Percent	Response Count
Strictly the ones with which the unit appears in my text.		72.3%	34
Most likely determiners or articles with which the unit can appear.		27.7%	13
answered question			47
skipped question			57

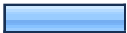





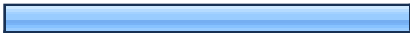


54. In what forms do you record your combinations of with determiner/adjective + unit?

		Response Percent	Response Count
In their base form		61.7%	29
In the form they appear in the text		42.6%	20
In all gender forms		2.1%	1
In all number forms		2.1%	1
In all case forms		2.1%	1
		answered question	47
		skipped question	57







55. Do you store multiple forms of combinations in one single record or in separate records?

		Response Percent	Response Count
Each combination in a separate record		61.7%	29
All combinations in a single record		38.3%	18
		answered question	47
		skipped question	57








56. Do you use any complimentary tool to identify potential collocations or forms of an expression? (Please, select at least one option to be properly directed to the next page.)

		Response Percent	Response Count
No.		17.6%	9
Yes, paper dictionaries and/or thesauri.		51.0%	26
Yes, an electronic database of words' conceptual-semantic and lexical relations (e.g. WordNet, Visuwords, etc.).		13.7%	7
Yes, the search function in a word processor.		23.5%	12
Yes, the search function in a document management system.		7.8%	4
Yes, my TEnT search function.		47.1%	24
Yes, an Internet search engine.		60.8%	31
Yes, a syntactic analyzer/parser.		0.0%	0
Yes, a concordancer that works on my documents.		27.5%	14
Yes, a concordancer that works on the WWW.		15.7%	8
Other (please specify)		0.0%	0
answered question			51
skipped question			53






57. Why do you not use the Terminology Management System integrated with your TEnT?

		Response Percent	Response Count
I already had developed my terminology database in another system.		9.5%	2
My TEnT Terminology Management System does not meets my needs.		14.3%	3
My TEnT Terminology Management System is too complex.		28.6%	6
Another TMS or tool meets my needs better.		19.0%	4
I never learnt how to use it.		23.8%	5
Other (please specify)		28.6%	6
		answered question	21
		skipped question	83



58. What tool do you use to store your terminology information?

		Response Percent	Response Count
Notebook, index cards or any other paper-format approach		4.5%	1
Word processor		9.1%	2
Spreadsheet		22.7%	5
General database (e.g. Access)		13.6%	3
Stand-alone off-the-shelf Terminology Management System		9.1%	2
In-house Terminology Management System		18.2%	4
Another Terminology Management System integrated with another TEnT		0.0%	0
Other (please specify)		22.7%	5
		answered question	22
		skipped question	82


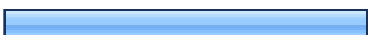
59. Why do you not record any terminology?

		Response Percent	Response Count
I consult existing terminological resources (e.g. paper dictionaries and glossaries, on-line terminology databases and glossaries).		75.0%	3
I consult existing on-line corpora.		25.0%	1
I consult the WWW.		100.0%	4
All terminology information can be found in my repository of past translations (e.g. archive, translation memory, corpora, etc.).		75.0%	3
I do not have the time.		0.0%	0
I have not found a tool that suits my terminological needs.		0.0%	0
I do not bill for my terminological work.		0.0%	0
I do not think it has a value.		0.0%	0
I know all the terminology I need.		0.0%	0
I do not know how.		0.0%	0
I don't know.		0.0%	0
Other (please specify)		25.0%	1
answered question			4
skipped question			100


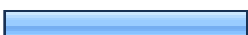
60. Do you think having additional features would encourage you to start recording terminology? If so, what kind?

		Response Percent	Response Count
No		45.5%	5
Yes (please specify)		54.5%	6
answered question			11
skipped question			93

61. Do you think receiving (additional) training would encourage you to start recording terminology? If so, what kind?

		Response Percent	Response Count
No		45.5%	5
Yes (please specify)		54.5%	6
answered question			11
skipped question			93




62. Do you think having access to more or different kinds of documentation would encourage you to start recording terminology? If so, what kind?

		Response Percent	Response Count
No		63.6%	7
Yes (please specify)		36.4%	4
answered question			11
skipped question			93




63. Is there anything else that could encourage you to record terminology?

	Response Count
	2
answered question	2
skipped question	102

64. Do you plan to record terminology information in the future?

		Response Percent	Response Count
Yes		40.0%	4
No		20.0%	2
Don't know		40.0%	4
	answered question		10
	skipped question		94

65. What tool do you see yourself using to record terminology in the future?

		Response Percent	Response Count
Notebook, index cards or any other paper-format approach		0.0%	0
Word processor		33.3%	2
Spreadsheet		16.7%	1
General database (e.g. Access)		0.0%	0
Stand-alone off-the-shelf Terminology Management System		0.0%	0
In-house Terminology Management System		0.0%	0
Terminology Management System integrated with a TEnT		50.0%	3
answered question			6
skipped question			98

66. During this survey, we mentioned some strategies to enhance your terminology information retrieval capacity. Do you use any similar strategy we did not mention in the survey? If so, please describe it below.

	Response Count
	6
answered question	6
skipped question	98

67. If we have mentioned the strategies you apply, but missed tackling any aspect you consider relevant, please describe these aspects in the text box below.

	Response Count
	3
answered question	3
skipped question	101

68. The results of this survey will reveal the actual use the members of the language industry make of Terminology Management Systems and it may also open new doors to explore.

Would you be interested in receiving any subsequent surveys on this topic? If so, please enter your e-mail address in the text box below.



	Response Count
	43
answered question	43
skipped question	61

69. If there is any information you want to share and felt the questions did not let you express it clearly, or if you have any additional comments to add, please, feel free to use the text box below.


	Response Count
	13
answered question	13
skipped question	91

APPENDIX F: Integrated Termbases Optimization Survey Results

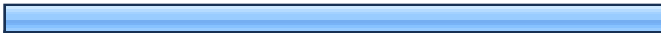

1. Do you agree to participate in the above research study conducted by Marta Gómez Palou of the School of Translation and Interpretation – Faculty of Arts (University of Ottawa) which research is under the supervision of Dr. Lynne Bowker and Dr. Elizabeth Marshman?

		Response Percent	Response Count
Yes, I agree.		99.2%	121
No, I do not agree.		0.8%	1
answered question			122
skipped question			0



2. Are you 18-years old or older?

		Response Percent	Response Count
Yes		100.0%	121
No		0.0%	0
answered question			121
skipped question			1



3. This survey is only available in English. Do you have good reading comprehension in English?

		Response Percent	Response Count
Yes		99.2%	120
No		0.8%	1
answered question			121
skipped question			1



4. Do you use a TEnT such as SDL Trados, Déjà Vu, MultiTrans, LogiTerm, Wordfast, MemoQ, Omega-T or a similar tool?

		Response Percent	Response Count
Yes		90.8%	109
No		9.2%	11
answered question			120
skipped question			2



5. Leaving aside the possibility of other synonyms existing for this particular example, when translating interactively a sentence containing a term such as "aspirin", would you prefer the TEnT to propose the results as displayed in Option A or in Option B?

		Response Percent	Response Count
Option A		74.1%	80
Option B		25.9%	28
answered question			108
skipped question			14



6. When translating interactively a sentence containing an acronym, would you prefer the TEnT to proposed the results as displayed in Option A or in Option B?

		Response Percent	Response Count
Option A		89.0%	97
Option B		11.0%	12
answered question			109
skipped question			13



7. When pretranslating a sentence containing a term with multiple equivalents, would you prefer the TEnT to propose the results as displayed in Option A or in Option B?

		Response Percent	Response Count
Option A		36.1%	39
Option B		63.9%	69
answered question			108
skipped question			14



8. Would you consider it useful to create a term record in your TEnT integrated termbase for an email address that changes in the target language?

		Response Percent	Response Count
Yes		61.1%	66
No		38.9%	42
answered question			108
skipped question			14



9. Would you consider it useful to create a term record in your TEnT integrated termbase for a civic address that changes in the target language?

		Response Percent	Response Count
Yes		68.8%	75
No		31.2%	34
answered question			109
skipped question			13



10. Would you consider it useful to create a term record in your TEnT integrated termbase for a URL that changes in the target language?

		Response Percent	Response Count
Yes		62.0%	67
No		38.0%	41
answered question			108
skipped question			14




11. Would you consider it useful to create a term record in your TEnT integrated termbase for a standard text that appears frequently and that must be translated by a standard target translation?

		Response Percent	Response Count
Yes		69.7%	76
No		30.3%	33
answered question			109
skipped question			13



12. When creating the record for this unit, what form or forms would you add to your termbase?

		Response Percent	Response Count
Only the base form. In this case, as the unit is a verb, the infinitive form: <i>marinate</i>		58.7%	64
Multiple forms. In this case, it could be all or any combination or of the following forms: <i>marinate</i> , <i>marinating</i> or <i>marinated</i> .		41.3%	45
		answered question	109
		skipped question	13





13. In the last question you indicated that you would record multiple forms of the unit we are considering. Which forms would you record?

		Response Percent	Response Count
All the forms of the unit that appear in the text. In this example: <i>marinate</i> , <i>marinating</i> and <i>marinated</i> .		37.8%	17
The most frequent form that appears in the text. In this example: <i>marinating</i> .		8.9%	4
All the forms of the unit that appear in the text and any other form I can think of. In this example: <i>marinate</i> , <i>marinating</i> , <i>marinated</i> and <i>marinates</i> .		53.3%	24
		answered question	45
		skipped question	77







14. Consider the structure (available fields and options, mandatory fields, limitations, etc.) of the TBX-Basic record template. If you started to use a new TEnT which had the option to use a built-in term record template based on TBX-Basic, where the only modifications allowed would be renaming fields and deleting non-mandatory fields, would you choose to work with this TBX-compliant template?

		Response Percent	Response Count
Yes, I would base my termbase on the TBX-Basic term record template and would work within its conditions.		72.2%	78
No, I would create my own term record template.		27.8%	30
		answered question	108
		skipped question	14


15. You indicated you would use a term record template based on TBX-Basic. Select the reason below that most influenced your decision.

		Response Percent	Response Count
The record template is based on an industry-approved standard and compliant with an international standard.		13.8%	11
It is easier than creating my own record template from scratch.		18.8%	15
Using a TBX-Basic template ensures that I will be able to easily share my termbase in the future and import other TBX-compliant termbases.		42.5%	34
The TBX-Basic template meets my terminological needs.		25.0%	20
Other (please specify)		0.0%	0
		answered question	80
		skipped question	42

16. You indicated you would prefer using your own term record template over a TBX-Basic template. Select the reason below that most influenced your decision.




		Response Percent	Response Count
TBX-Basic does not include a field that I require in my termbase.		19.4%	6
I do not agree with the term field being mandatory.		3.2%	1
I do not agree with the language field being mandatory.		0.0%	0
I do not agree with either a part of speech, a definition or a context being mandatory.		25.8%	8
TBX-Basic does not include pick list values that I require.		16.1%	5
I am required to use a specific record template structure to maintain compatibility to exchange data with a certain software, client or institution.		9.7%	3
Other		25.8%	8
		answered question	31
		skipped question	91

17. Country of residence:


		Response Percent	Response Count
Canada		22.4%	24
France		4.7%	5
Spain		4.7%	5
Germany		6.5%	7
-----		0.0%	0
Afghanistan		0.0%	0
Albania		0.0%	0
Algeria		0.0%	0
Andorra		0.0%	0
Angola		0.0%	0
Antigua & Deps		0.0%	0
Argentina		4.7%	5
Armenia		0.0%	0
Australia		0.0%	0
Austria		3.7%	4
Azerbaijan		0.0%	0
Bahamas		0.0%	0
Bahrain		0.0%	0
Bangladesh		0.0%	0
Barbados		0.0%	0
Belarus		0.0%	0
Belgium		2.8%	3
Belize		0.0%	0



Benin		0.0%	0
Bhutan		0.0%	0
Bolivia		0.0%	0
Bosnia Herzegovina		0.0%	0
Botswana		0.0%	0
Brazil		1.9%	2
Brunei		0.0%	0
Bulgaria		0.0%	0
Burkina		0.0%	0
Burundi		0.0%	0
Cambodia		0.0%	0
Cameroon		0.0%	0
Cape Verde		0.0%	0
Central African Rep		0.0%	0
Chad		0.0%	0
Chile		0.9%	1
China		0.9%	1
Colombia		0.0%	0
Comoros		0.0%	0
Congo		0.0%	0
Congo {Democratic Rep}		0.0%	0
Costa Rica		0.0%	0
Croatia		0.0%	0
Cuba		0.0%	0
Cyprus		0.0%	0





Czech Republic		1.9%	2
Denmark		1.9%	2
Djibouti		0.0%	0
Dominica		0.0%	0
Dominican Republic		0.0%	0
East Timor		0.0%	0
Ecuador		0.9%	1
Egypt		0.0%	0
El Salvador		0.0%	0
Equatorial Guinea		0.0%	0
Eritrea		0.0%	0
Estonia		0.0%	0
Ethiopia		0.0%	0
Fiji		0.0%	0
Finland		0.9%	1
Gabon		0.0%	0
Gambia		0.0%	0
Georgia		0.0%	0
Ghana		0.0%	0
Greece		0.9%	1
Grenada		0.0%	0
Guatemala		0.0%	0
Guinea		0.0%	0
Guinea-Bissau		0.0%	0
Guyana		0.0%	0
Haiti		0.0%	0

Honduras		0.0%	0
Hungary		0.0%	0
Iceland		0.9%	1
India		0.0%	0
Indonesia		0.9%	1
Iran		0.0%	0
Iraq		0.0%	0
Ireland {Republic}		0.0%	0
Israel		0.0%	0
Italy		2.8%	3
Ivory Coast		0.0%	0
Jamaica		0.0%	0
Japan		0.9%	1
Jordan		0.0%	0
Kazakhstan		0.0%	0
Kenya		0.0%	0
Kiribati		0.0%	0
Korea North		0.0%	0
Korea South		0.0%	0
Kosovo		0.0%	0
Kuwait		0.0%	0
Kyrgyzstan		0.0%	0
Laos		0.0%	0
Latvia		0.0%	0
Lebanon		0.0%	0
Lesotho		0.0%	0





Liberia	0.0%	0
Libya	0.0%	0
Liechtenstein	0.0%	0
Lithuania	0.0%	0
Luxembourg	0.0%	0
Macedonia	0.0%	0
Madagascar	0.0%	0
Malawi	0.0%	0
Malaysia	0.0%	0
Maldives	0.0%	0
Mali	0.0%	0
Malta	0.0%	0
Marshall Islands	0.0%	0
Mauritania	0.0%	0
Mauritius	0.0%	0
Mexico	0.0%	0
Micronesia	0.0%	0
Moldova	0.0%	0
Monaco	0.0%	0
Mongolia	0.0%	0
Montenegro	0.0%	0
Morocco	0.0%	0
Mozambique	0.0%	0
Myanmar, {Burma}	0.0%	0
Namibia	0.0%	0

Nauru		0.0%	0
Nepal		0.0%	0
Netherlands		0.0%	0
New Zealand		0.0%	0
Nicaragua		0.0%	0
Niger		0.0%	0
Nigeria		0.0%	0
Norway		0.9%	1
Oman		0.0%	0
Pakistan		0.0%	0
Palau		0.0%	0
Panama		0.0%	0
Papua New Guinea		0.0%	0
Paraguay		0.0%	0
Peru		0.0%	0
Philippines		0.0%	0
Poland		2.8%	3
Portugal		1.9%	2
Qatar		0.0%	0
Romania		0.0%	0
Russian Federation		0.0%	0
Rwanda		0.0%	0
St Kitts & Nevis		0.0%	0
St Lucia		0.0%	0
Saint Vincent & the Grenadines		0.0%	0
Samoa		0.0%	0



San Marino		0.0%	0
Sao Tome & Principe		0.0%	0
Saudi Arabia		0.0%	0
Senegal		0.0%	0
Serbia		0.0%	0
Seychelles		0.0%	0
Sierra Leone		0.0%	0
Singapore		0.0%	0
Slovakia		0.0%	0
Slovenia		0.9%	1
Solomon Islands		0.0%	0
Somalia		0.0%	0
South Africa		0.0%	0
Sri Lanka		0.0%	0
Sudan		0.0%	0
Suriname		0.0%	0
Swaziland		0.0%	0
Sweden		0.0%	0
Switzerland		10.3%	11
Syria		0.0%	0
Taiwan		0.0%	0
Tajikistan		0.0%	0
Tanzania		0.0%	0
Thailand		0.0%	0
Togo		0.0%	0
Tonga		0.0%	0

Trinidad & Tobago		0.0%	0
Tunisia		0.0%	0
Turkey		0.9%	1
Turkmenistan		0.0%	0
Tuvalu		0.0%	0
Uganda		0.0%	0
Ukraine		0.9%	1
United Arab Emirates		0.0%	0
United Kingdom		2.8%	3
United States		11.2%	12
Uruguay		1.9%	2
Uzbekistan		0.0%	0
Vanuatu		0.0%	0
Vatican City		0.0%	0
Venezuela		0.0%	0
Vietnam		0.0%	0
Yemen		0.0%	0
Zambia		0.0%	0
Zimbabwe		0.0%	0
answered question			107
skipped question			15




18. Profession:





		Response Percent	Response Count
Translator		68.8%	75
Terminologist		7.3%	8
Project Manager		6.4%	7
Other (please specify)		17.4%	19
answered question			109
skipped question			13

19. Work setting:

		Response Percent	Response Count
Freelancer		56.4%	57
In-house translation service		43.6%	44
answered question			101
skipped question			21

20. Main source language:






		Response Percent	Response Count
English		53.3%	57
-----		0.0%	0
Catalan		0.0%	0
French		9.3%	10
German		18.7%	20
Spanish		2.8%	3
-----		0.0%	0
Afrikaans		0.0%	0
Albanian		0.0%	0
Amharic		0.0%	0
Arabic		0.0%	0
Aragonese		0.0%	0
Armenian		0.0%	0
Asturian		0.0%	0
Asturleonese		0.0%	0
Aymara		0.0%	0
Azerbaijani		0.0%	0
Bable		0.0%	0
Basque		0.0%	0
Belarusian		0.0%	0
Bengali		0.0%	0
Bosnian		0.0%	0
Breton		0.0%	0

Bulgarian		0.0%	0
Burmese		0.0%	0
Central Khmer		0.0%	0
Chinese		0.0%	0
Croatian		0.0%	0
Czech		0.0%	0
Danish		1.9%	2
Dutch		1.9%	2
English		5.6%	6
Estonian		0.0%	0
Filipino		0.0%	0
Finnish		0.9%	1
Flemish		0.0%	0
Gaelic		0.0%	0
Galician		0.0%	0
Georgian		0.0%	0
Greek, Modern (1453-)		0.0%	0
Greenlandic		0.0%	0
Guarani		0.0%	0
Hawaiian		0.0%	0
Hebrew		0.0%	0
Hindi		0.0%	0
Hungarian		0.0%	0
Icelandic		0.0%	0
Indonesian		0.0%	0



Inuktitut		0.0%	0
Irish		0.0%	0
Italian		3.7%	4
Japanese		0.9%	1
Korean		0.0%	0
Kurdish		0.0%	0
Latvian		0.0%	0
Lithuanian		0.0%	0
Macedonian		0.0%	0
Malay		0.0%	0
Maltese		0.0%	0
Masai		0.0%	0
Mi'kmaq		0.0%	0
Moldavian		0.0%	0
Moldovan		0.0%	0
Mong		0.0%	0
Mongolian		0.0%	0
Nepali		0.0%	0
Norwegian Bokmål		0.9%	1
Norwegian Nynorsk		0.0%	0
Polish		0.0%	0
Portuguese		0.0%	0
Punjabi		0.0%	0
Pushto		0.0%	0
Quechua		0.0%	0
Romanian		0.0%	0

Romansh	0.0%	0
Russian	0.0%	0
Serbian	0.0%	0
Slovak	0.0%	0
Slovenian	0.0%	0
Somali	0.0%	0
Swahili	0.0%	0
Swedish	0.0%	0
Tagalog	0.0%	0
Tahitian	0.0%	0
Tajik	0.0%	0
Tamil	0.0%	0
Thai	0.0%	0
Tibetan	0.0%	0
Turkish	0.0%	0
Ukrainian	0.0%	0
Urdu	0.0%	0
Vietnamese	0.0%	0
Welsh	0.0%	0
Yiddish	0.0%	0
Zulu	0.0%	0
answered question		107
skipped question		15

21. Main target language:

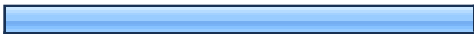

		Response Percent	Response Count
English		23.4%	25
-----		0.0%	0
Catalan		2.8%	3
French		28.0%	30
German		11.2%	12
Spanish		12.1%	13
-----		0.0%	0
Afrikaans		0.0%	0
Albanian		0.0%	0
Amharic		0.0%	0
Arabic		0.0%	0
Aragonese		0.0%	0
Armenian		0.0%	0
Asturian		0.0%	0
Asturleonese		0.0%	0
Aymara		0.0%	0
Azerbaijani		0.0%	0
Bable		0.0%	0
Basque		0.0%	0
Belarusian		0.0%	0
Bengali		0.0%	0
Bosnian		0.0%	0
Breton		0.0%	0

Bulgarian		0.0%	0
Burmese		0.0%	0
Central Khmer		0.0%	0
Chinese	▮	0.9%	1
Croatian		0.0%	0
Czech	▮	1.9%	2
Danish		0.0%	0
Dutch	▮	0.9%	1
English	▮	0.9%	1
Estonian		0.0%	0
Filipino		0.0%	0
Finnish		0.0%	0
Flemish		0.0%	0
Gaelic		0.0%	0
Galician		0.0%	0
Georgian		0.0%	0
Greek, Modern (1453-)	▮	0.9%	1
Greenlandic		0.0%	0
Guarani		0.0%	0
Hawaiian		0.0%	0
Hebrew		0.0%	0
Hindi		0.0%	0
Hungarian		0.0%	0
Icelandic	▮	0.9%	1
Indonesian		0.0%	0

Inuktitut		0.0%	0
Irish		0.0%	0
Italian		3.7%	4
Japanese		0.0%	0
Korean		0.0%	0
Kurdish		0.0%	0
Latvian		0.0%	0
Lithuanian		0.0%	0
Macedonian		0.0%	0
Malay		0.0%	0
Maltese		0.0%	0
Masai		0.0%	0
Mi'kmaq		0.0%	0
Moldavian		0.0%	0
Moldovan		0.0%	0
Mong		0.0%	0
Mongolian		0.0%	0
Nepali		0.0%	0
Norwegian Bokmål		0.0%	0
Norwegian Nynorsk		0.0%	0
Polish		3.7%	4
Portuguese		4.7%	5
Punjabi		0.0%	0
Pushto		0.0%	0
Quechua		0.0%	0
Romanian		0.0%	0

Romansh		0.0%	0
Russian	▮	0.9%	1
Serbian		0.0%	0
Slovak		0.0%	0
Slovenian	▮	0.9%	1
Somali		0.0%	0
Swahili		0.0%	0
Swedish		0.0%	0
Tagalog		0.0%	0
Tahitian		0.0%	0
Tajik		0.0%	0
Tamil		0.0%	0
Thai		0.0%	0
Tibetan		0.0%	0
Turkish	▮	0.9%	1
Ukrainian	▮	0.9%	1
Urdu		0.0%	0
Vietnamese		0.0%	0
Welsh		0.0%	0
Yiddish		0.0%	0
Zulu		0.0%	0
answered question			107
skipped question			15

22. Have you received any formal education (university courses, certificate, workshop) on terminology theory and/or practices?

		Response Percent	Response Count
Yes		70.6%	77
No		29.4%	32
		answered question	109
		skipped question	13