



uOttawa

L'Université canadienne
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES**



**FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES**

Cheryl McBride

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Translation)

GRADE / DEGREE

School of Translation and Interpretation

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Translation Memory System: An Analysis of Translator's Attitudes and Opinions

TITRE DE LA THÈSE / TITLE OF THESIS

Lynne Bowker

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Elizabeth Marshman

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Debbie Folaron

Jean Quirion

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Translation Memory Systems: An Analysis of Translators' Attitudes and Opinions

Cheryl McBride

School of Translation and Interpretation
University of Ottawa

Under the supervision of
Lynne Bowker, PhD and
Elizabeth Marshman, PhD
School of Translation and Interpretation
University of Ottawa

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of MA (Translation)



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-61311-5
Our file Notre référence
ISBN: 978-0-494-61311-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ♦ ■
Canada

Abstract

Translation memory (TM) systems are among the most aggressively marketed and widely used computer-aided translation tools. Previous studies have focused on when and how TMs are used, but there is significantly less information available relating to translators' perceptions of and attitudes towards them.

The goal of this thesis is to explore translators' unprompted opinions of the issues related to TM system usage. After analyzing postings on translators' discussion boards, I propose to compare current assumptions about TM systems and their use with what translators are expressing in their unprompted opinions. I believe that with a better understanding of different perspectives and attitudes, translators can evaluate and potentially adjust their own perceptions in light of others' experience, developers and vendors can respond more accurately to users' needs, clients can better comprehend translators' concerns, and researchers and trainers can properly address the issues currently surrounding TM system usage.

This thesis is organized into three chapters. Following a general introduction, Chapter 1 explains the functioning of TM systems and the issues surrounding their use, and then explores what is known about the use of TM systems and attitudes towards them as these are expressed in scholarly research, vendor promotional materials, surveys of practicing translators, and analyses of mailing lists. Chapter 2 provides a description of the methodology used in this project to select a primary resource, extract TM-related information, and classify the data. Chapter 3 presents a summary and analysis of the data found in the corpus. Finally, the conclusion summarizes the findings of this research and their implications for translators, vendors, clients/agencies, translator trainers, and researchers, addresses areas requiring further investigation and research, and evaluates the methodology of the project.

Résumé

Les mémoires de traduction (MT) sont parmi les outils de traduction humaine assistée par ordinateur les plus vigoureusement commercialisés et les plus utilisés. Les études jusqu'à maintenant ont porté sur quand et comment les MT sont utilisées, mais il y a beaucoup moins d'information sur les perceptions de ces outils par les traducteurs et les attitudes de ces derniers envers les MT.

L'objectif de la présente thèse est d'explorer les opinions a priori des traducteurs au sujet des questions liées à l'utilisation des MT. Après avoir analysé des billets sur des forums de discussion de traducteurs, je propose de comparer les idées reçues sur les MT et leur utilisation avec ce que disent les traducteurs. Je crois qu'en comprenant mieux les différentes perspectives et attitudes, les traducteurs pourront évaluer et potentiellement ajuster leurs perceptions en fonction des expériences des autres; les concepteurs et vendeurs de MT pourront mieux répondre aux besoins des utilisateurs; les clients pourront mieux comprendre les inquiétudes des traducteurs et les chercheurs et formateurs pourront bien cerner les problèmes de l'utilisation des MT.

La présente thèse est composée de trois chapitres. L'introduction générale est suivie du Chapitre 1, qui explique le fonctionnement des MT et les questions sur leur utilisation, et explore les idées reçues sur l'utilisation des MT qui se trouvent dans les recherches universitaires, les dépliants publicitaires des vendeurs de MT, les sondages de traducteurs en exercice et les analyses de listes d'envoi. Le deuxième chapitre décrit la méthodologie employée pour choisir une ressource principale, extraire des informations liées aux MT et trier les données. Le troisième chapitre résume et présente l'analyse des données extraites du corpus. Enfin, la conclusion englobe un sommaire des résultats et les conséquences pour les traducteurs, les vendeurs, les clients et cabinets de traduction, les formateurs de traducteurs et les chercheurs; elle traite également des domaines qui devraient être examinés plus à fond et évalue la méthodologie du projet.

Acknowledgements

I would like to thank Lynne Bowker and Elizabeth Marshman for their never-ending patience and unfailing support. You helped me keep my sanity *and* verbosity in check. This thesis would not be what it is if not for your willingness to read innumerable pages, and provide encouraging and insightful feedback.

To my family – thank you. Yes, Dad, I’m finally finished. You were all there to encourage and listen, and I would not be where I am without your confidence in me and my abilities. You stuck with me through the Master’s, now on to the PhD!

My dearest fellow thesis writers: We’ve come a long way, and it was my pleasure to share this experience with you. We made it through the ups and downs, the moments of laughter and the cries of desperation. I am grateful to have such a supportive and caring group of friends!

I offer my sincerest apologies to all who had to listen to me talk, complain, and whine about my corpus (and my thesis in general). I promise to never create a corpus again...well, at least not for a little while.

Finally, I would like to acknowledge the members and staff of ProZ.com and TranslatorsCafé.com. You have created a friendly and safe environment in which translators can freely share their thoughts and opinions, and receive advice and help from their peers. The forums on these sites are invaluable, and I would have no thesis without your contributions.

Table of Contents

ABSTRACT.....	II
RÉSUMÉ	III
ACKNOWLEDGEMENTS.....	IV
LIST OF TABLES.....	X
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS.....	XII
INTRODUCTION.....	1
0.1 Context and motivation.....	1
0.2 Objectives	3
0.3 Scope and limitations.....	3
0.4 General methodology.....	4
0.5 Outline.....	5
CHAPTER 1 – THE TRANSLATION MEMORY SYSTEM	6
1.1 Translation memories: Inception to adoption	6
1.1.1 The Birth of TM systems	6
1.1.2 TM systems today	7
1.2 Components of translation environments	8
1.2.1 TM systems.....	8
1.2.2 Other TEnT components.....	9
1.3 Purchasing or acquiring a TM system	10
1.3.1 Situational considerations	10
1.3.1.1 Text characteristics and format	10
1.3.1.2 Number of users	12
1.3.1.3 Languages.....	12
1.3.1.3.1 Language pairs.....	12
1.3.1.3.2 Language of the interface	13
1.3.2 Technical considerations.....	14
1.3.2.1 Computing requirements	14
1.3.2.1.1 Processing speed and memory	14
1.3.2.1.2 Operating system	14
1.3.2.2 Open-source versus closed-source software.....	15
1.3.2.3 Word-processor or text-editor integration	15
1.3.2.4 Voice recognition software.....	16
1.3.2.5 TM creation, storage, and usage methods	16
1.3.2.5.1 Sentence-based and character-string-within-a-bitext approaches.....	16
1.3.2.5.2 Interactive creation and post-translation alignment.....	18
1.3.2.5.3 Pretranslation and interactive translation modes	19
1.3.3 Brand-specific considerations.....	20
1.3.3.1 Trial or demo versions.....	20
1.3.3.2 Price.....	20
1.3.3.3 Updates or new releases	21
1.4 Using a TM system	22
1.4.1 Training.....	22
1.4.1.1 Time	22
1.4.1.2 Availability.....	23
1.4.1.3 Price.....	24
1.4.2 File formats	24
1.4.2.1 Input and output files.....	25
1.4.2.1.1 Source and target file formats	25

1.4.2.1.2 Legacy file formats.....	26
1.4.2.1.3 Files created by the TM system	26
1.4.2.1.4 TM formats.....	27
1.4.2.2 Compatibility among TM systems	27
1.4.2.3 Vendor lock-in.....	28
1.4.2.4 Open standard formats.....	28
1.4.3 Matches proposed by a TM system	29
1.4.3.1 Types of matches.....	30
1.4.3.2 Usefulness of matches	32
1.4.4 Technical issues	33
1.4.4.1 Customizability of TM systems.....	33
1.4.4.2 Technical support	34
1.4.5 Procedural and textual effects of TM usage.....	35
1.4.5.1 Speed.....	35
1.4.5.2 Quality.....	35
1.4.6 Client relations	36
1.4.6.1 Brand or format requirements	36
1.4.6.2 Payment.....	37
1.4.6.3 Ownership	38
1.4.6.4 Buying/selling/trading TMs.....	38
1.5 Vendor promotional materials	40
1.6 TM-related surveys	42
1.6.1 Discussion of survey methodologies	45
1.6.1.1 Respondents	45
1.6.1.2 Areas of interest	45
1.6.2 Survey summaries.....	46
1.6.2.1 LISA 2002	46
1.6.2.2 LISA 2004.....	47
1.6.2.3 eCoLoRe.....	51
1.6.2.4 Fulford and Granell-Zafra	52
1.6.2.5 Dillon and Fraser	53
1.6.2.6 Lagoudaki.....	54
1.6.2.7 ATIO 2005	57
1.6.2.8 ATIO 2007	57
1.6.2.9 OTTIAQ.....	58
1.6.3 Comparison of TM surveys	59
1.6.3.1 TM usage.....	59
1.6.3.2 Tool brands.....	60
1.6.3.3 Non-adoption of TM systems.....	62
1.6.3.4 Training	62
1.6.3.5 New developments in TM technology.....	63
1.7 García and Lantra-L	63
1.8 Concluding remarks	65
CHAPTER 2 – METHODOLOGY	67
2.1 Choice of primary source.....	67
2.1.1 TM user interviews	67
2.1.2 TM usage survey.....	67
2.1.3 Mailing lists	68
2.1.4 Discussion boards	69
2.1.4.1 Privacy and discussion boards.....	70
2.1.4.2 Terminological clarifications.....	71
2.2 Discussion board selection.....	75
2.2.1 English Spanish Translator Org	75

2.2.2	Yahoo!	76
2.2.3	Global Languages	76
2.2.4	aquarius.net	77
2.2.5	TranslatorsCafé.com	77
2.2.6	ProZ.com	77
2.3	Sample corpus	78
2.4	Forum selection	81
2.4.1	Board-wide keyword search	82
2.4.2	Creation of word list	82
2.4.3	Selection of keywords	82
2.4.4	<i>A priori</i> exclusions	83
2.4.5	ProZ.com forum search	83
2.4.6	TranslatorsCafé.com forum search	84
2.4.7	<i>A posteriori</i> exclusions	84
2.4.8	Forum search summary	85
2.5	Thread selection	85
2.6	Collection of postings	86
2.7	Corpus structure	87
2.7.1	Administrative information	87
2.7.2	Content-related information	88
2.7.3	Classification system	88
2.7.3.1	Topic-related categories	89
2.7.3.2	Brand-name categories	90
2.7.3.3	Operating-system categories	91
2.8	Concluding remarks	92
CHAPTER 3 – CORPUS ANALYSIS		93
3.1	Corpus statistics	93
3.1.1	Topic-related category statistics	94
3.1.2	Brand-name category statistics	98
3.1.3	Operating-system category statistics	100
3.2	Analysis and summary of corpus content	100
3.2.1	Purchasing or acquiring a TEnT	101
3.2.1.1	Situational considerations	101
3.2.1.1.1	Text characteristics and format	101
3.2.1.1.2	Number of users	104
3.2.1.1.3	Languages	104
3.2.1.1.3.1	Language pairs	104
3.2.1.1.3.2	Language of the interface	106
3.2.1.2	Technical considerations	106
3.2.1.2.1	Computing requirements	106
3.2.1.2.1.1	Processing speed and memory	106
3.2.1.2.1.2	Operating system	107
3.2.1.2.2	Open-source versus closed-source software	108
3.2.1.2.3	Word-processor or text-editor integration	109
3.2.1.2.4	Voice recognition software	110
3.2.1.2.5	TM creation, storage, and usage methods	110
3.2.1.3	Brand-specific considerations	111
3.2.1.3.1	Trial or demo versions	111
3.2.1.3.2	Price	111
3.2.1.3.3	Updates or new releases	114
3.2.2	Using a TM system	116
3.2.2.1	Training	116

3.2.2.1.1	Time.....	117
3.2.2.1.2	Availability.....	118
3.2.2.1.3	Price.....	120
3.2.2.2	File formats	120
3.2.2.2.1	Input and output files.....	120
3.2.2.2.1.1	Source and target file formats	121
3.2.2.2.1.2	Legacy file formats	122
3.2.2.2.1.3	Files created by the TM system	122
3.2.2.2.1.4	TM formats	122
3.2.2.2.2	Compatibility among TM systems.....	122
3.2.2.2.3	Vendor lock-in.....	124
3.2.2.2.4	Open standard formats.....	126
3.2.2.3	Matches proposed by a TM system	127
3.2.2.3.1	Types of matches	127
3.2.2.3.2	Usefulness of matches	127
3.2.2.4	Technical issues.....	128
3.2.2.4.1	Customizability of TM systems.....	128
3.2.2.4.2	Technical support	129
3.2.2.5	Procedural and textual effects of TM usage	132
3.2.2.5.1	Speed	132
3.2.2.5.2	Quality.....	132
3.2.2.5.3	Modifications to the translation process	133
3.2.2.6	Client relations	134
3.2.2.6.1	Brand or format requirements.....	134
3.2.2.6.2	Payment.....	136
3.2.2.6.3	Ownership.....	137
3.2.2.6.4	Buying/selling/trading TMs.....	138
3.2.2.7	Non-adoption attitudes	138
3.2.3	Comparison with vendor promotional materials.....	139
3.3	Concluding remarks	141
CONCLUSION		142
4.1	Observations about contributing authors	142
4.2	Interpretation of corpus contents	144
4.2.1	Purchasing or acquiring a TM system.....	144
4.2.1.1	Situational considerations.....	144
4.2.1.1.1	Text characteristics and format.....	144
4.2.1.1.2	Number of users.....	145
4.2.1.1.3	Languages.....	146
4.2.1.2	Technical considerations	146
4.2.1.2.1	Computing requirements	146
4.2.1.2.2	Open-source versus closed-source software	147
4.2.1.2.3	Word-processor or text-editor integration	148
4.2.1.2.4	Voice recognition software.....	148
4.2.1.2.5	TM creation, storage, and usage methods.....	149
4.2.1.3	Brand-specific considerations	150
4.2.1.3.1	Trial or demo versions	150
4.2.1.3.2	Price.....	150
4.2.1.3.3	Updates or new releases.....	151
4.2.2	Using a TM system	152
4.2.2.1	Training	152
4.2.2.2	File formats	153
4.2.2.3	Matches proposed by a TM system	154
4.2.2.4	Technical issues.....	155
4.2.2.5	Procedural and textual effects of TM usage	156
4.2.2.6	Client relations	157

4.3 Evaluation of the methodology	159
4.4 Concluding remarks	162
REFERENCES.....	164
APPENDIX A.....	171
APPENDIX B.....	174
APPENDIX C.....	176
APPENDIX D.....	180
APPENDIX E.....	181
APPENDIX F.....	186
APPENDIX G.....	187
APPENDIX H.....	190
APPENDIX I.....	212

List of Tables

Table 1 - TEnT Components.....	9
Table 2 - Types of Matches Identified by a TM System	31
Table 3 - Top 20 TM-Related Promotional Points	41
Table 4 - Overview of TM Surveys	43
Table 5 - Survey Summaries.....	44
Table 6 - Ten Most Commonly Used Tools, LISA 2004	50
Table 7 - Ten Most Commonly Used Tools, Lagoudaki (2006).....	55
Table 8 - Overview of TM Usage	59
Table 9 - Comparison of Top Ten Tools, LISA 2004 and Lagoudaki (2006)	61
Table 10 - Forum Selection Results.....	85
Table 11 - Summary of Corpus Statistics	87
Table 12 - Hierarchical Listing of Topic-Related Categories.....	90
Table 13 - Hierarchical Listing of Brand-Name Categories	91
Table 14 – General Corpus Statistics.....	94
Table 15 - Topic-Related Category Statistics	96
Table 16 - Topic-Related Categories by Frequency	97
Table 17 - Brand-Name Category Statistics	99
Table 18 - Brand References Based on Tool Type	99
Table 19 - Comparison of Corpus, LISA 2004, and Lagoudaki Top Tools	100
Table 20 - Operating-System Category Statistics	100

List of Figures

Figure 1 - Discussion Board Organization..... 72

Figure 2 - Threads within a Forum 73

Figure 3 - Postings 74

Figure 4 - Sample Corpus 80

List of Abbreviations

ABI	Allied Business Intelligence
ALPAC	Automatic Language Processing Advisory Committee
ATIO	Association of Translators and Interpreters of Ontario
CAD	computer-aided design (or computer-aided drafting)
CAT	computer-aided translation
CMS	content management system
CSB	character-string-within-a-bitext
CTISC	Canadian Translation Industry Sectoral Committee
DTP	desktop publishing
EBMT	example-based machine translation
EC	European Commission
eCoLoRe	eContent Localization Resources
EU	European Union
FAQ	frequently asked question(s)
GILT	globalization, internationalization, localization, and translation
GUI	graphical user interface
HTML	hypertext markup language
ICT	information and communication technologies
IT	information technology
JRC	Joint Research Centre
LINFO	Linux Information Project
LISA	Localization Industry Standards Association
LSP	language service provider
MT	machine translation
OCR	optical character recognition
OS	operating system
OSCAR	Open Standards for Container/Content Allowing Re-use
OTTIAQ	Ordre des traducteurs, terminologues et interprètes agréés du Québec
ROI	return on investment
RTFM	read the friendly manual
SB	sentence-based
SRX	Segmentation Rules eXchange
TAUS	Translation Automation User Society
TBX	Term Base eXchange
TDA	TAUS Data Association
TEnT	translation environment
TM	translation memory
TMS	terminology management system
TMX	Translation Memory eXchange
TU	translation unit
URL	Uniform Resource Locator
VRS	voice recognition software
XLIFF	XML Localisation Interchange File Format
XML	extensible markup language

Introduction

Recent growth in the translation technologies sector has seen the emergence of translation memory (TM) systems as one of the most aggressively marketed and widely used specialized technologies for translators. While the range of technologies available for translators includes everything from electronic dictionaries and terminology extractors to highly sophisticated translation environments, TM systems in particular have transformed the translation profession as storage and retrieval capabilities have advanced at an extraordinary pace (Rode 2000, 12). Although the concept of the TM has existed for some time, it is only in recent years that this type of tool has become an important commercial product (Bowker 2002, 92). Numerous studies have focused on the quantitative use of TM systems (e.g. Lommel 2004; Fulford and Granell-Zafra 2005; Lagoudaki 2006); however, comparatively little research has been done on translators' attitudes towards these tools.¹ In this thesis, I aim to create a systematic framework for evaluating translators' attitudes towards TM systems and to perform an in-depth analysis of these unprompted opinions in order to offer insight into this previously under-investigated area.

0.1 Context and motivation

For the purpose of this project, I will use the term **translation memory (TM)** to refer to databases of documents and their translations in which the corresponding texts are aligned, allowing a user to quickly and easily see how a particular unit (i.e. sentence, paragraph, sub-segment chunk, text), which recurs in a new source text to be translated, was previously translated. The term **translation memory (TM) system** is used to refer to a computer program used to search these TM databases for matches. TM systems are intended to save translators time and effort, and to improve the consistency of translated documents by automatically identifying previously translated units and suggesting their translations to the user. Translating more quickly could in turn increase the translator's productivity, allowing him or her to meet shorter deadlines and earn more money. Furthermore, the integration of a TM system with numerous other tools in a translation

¹ Exceptions include Merkel (1998), García (2003) and (2006), and Lagoudaki (2006).

environment may allow the translator to take advantage of additional functions (e.g. terminology management, concordancing) during the translation process itself, rather than using these other tools before or after translating the text.

While previous studies have focused on when and how TM systems are used, there is significantly less information available relating to translators' perceptions of and attitudes towards these tools. Important questions that need to be answered include the following: Are translators discussing problems related to TM use and seeking help from their peers? What issues or concerns are being expressed? What kind of advice are translators looking for when purchasing a TM system? Are translators praising or criticizing certain TM systems or features? Are certain systems discussed more than others? What is the general level of user satisfaction among translators? Do translators feel that TM systems increase their productivity and consistency? Are TM systems only being used for certain types of texts? How have TMs affected the relationship between translators and clients?

The analysis of translators' opinions may provide some answers to these questions. One way of exploring attitudes is through the use of surveys; however, there is additional merit in studying these attitudes in unprompted and spontaneous expressions (e.g. in an online discussion forum). Moreover, this type of analysis can allow for the inclusion and examination of unpredicted or previously unaddressed opinions of translators (e.g. issues not addressed by questions included on a survey).

Such an analysis can help to increase awareness of translators' attitudes and assist in bridging the gap between end-users and developers of TM systems. Translators rarely have input in tool development, and, when they do, it is often in the late stages. Information such as that gained through this research and analysis may help commercial producers and employers to evaluate users' reactions and to present tools in a more effective and enticing way. It is also important for trainers in translator training programs to address both the functioning of TM systems and the issues surrounding their use in order to provide an accurate overview of the role of TM in the industry. Moreover, translators should recognize their peers as valuable sources of information and advice. Finally, clients and agencies need to better understand the concerns and needs of translators. By acknowledging and understanding these attitudes, translators, clients, and

vendors can improve their relationships with each other. With a better understanding of different perspectives and attitudes, translators can evaluate and potentially alter their own perceptions, producers of these tools can respond more accurately to their users' needs, clients can better comprehend the needs and concerns of translators, researchers can better reflect these attitudes and issues in the scholarly literature, and translator trainers can better prepare new graduates for the realities of the industry.

0.2 Objectives

This thesis aims to achieve a deeper understanding of translators' attitudes and perceptions of TM systems through the creation and analysis of a corpus containing these viewpoints and opinions. By examining translators' natural and unprompted reactions and evaluating how well current literature and vendor promotional materials reflect them, common issues and problems can be brought to the foreground and addressed through an expansion of the field's literature, further development of the tools by commercial producers, and improved training.

0.3 Scope and limitations

TM systems are not the only tools that have seen widespread use and development within the past few decades. Many other computer-aided translation (CAT) tools are available and are used by both in-house and freelance translators. These tools, and specifically TM systems, are not often used in a stand-alone fashion. Generally, numerous tools are integrated into a **translation environment (TE_nT)**, where they work together and interact with one another. TE_nTs can include tools such as TM systems, terminology management systems, term extractors, active terminology recognition, concordancers, and machine translation (MT) applications. However, in order to maintain a manageable scope, only information directly related to TM systems (i.e. the TM component of these environments) has been analyzed here.

Certain restrictions have been imposed when choosing a primary source: the source must contain sufficient and recent TM-related data in English. The primary source shall not be limited or restricted to a geographical area or language pairs. Finally,

the primary source must provide the views of users of TM systems, which can be accessed without focusing on particular brands or relying strictly on keyword searches.

The English restriction was imposed because English is my first language, and therefore is the language with which I am most comfortable. Furthermore, as one of the most widely spoken languages in the world, English is often the *lingua franca* used by a group of people with diverse linguistic backgrounds. The other criteria were established to ensure that numerous translators' opinions were included, and that the contents of the corpus did not reflect only one language-specific, brand-specific, or geographical reality of TM usage. Finally, the requirement for access without the use of keyword searches was established to reduce bias based on keyword choice (i.e. the keywords chosen may reduce the possibility of obtaining information on certain issues while increasing the possibility of obtaining information on other issues).

0.4 General methodology

First, a review of the available literature was performed to investigate what is currently known (or surmised) about the use of and attitudes towards TM systems. The review included trends in research and development of TM systems, anticipated benefits of use, potential drawbacks, vendor promotional materials, results from TM-related surveys of translators, and findings from previous investigations of translators opinions and attitudes towards TM usage, thus providing a framework of issues and topics previously and currently addressed by users, vendors, clients/agencies, trainers, and researchers.

Once the literature review was completed, a primary source and specific sources for data collection were identified. The structure of the corpus to be used for the analysis was defined, and the corpus was created. A classification system was designed to categorize the contents of the corpus and to facilitate analysis.

Finally, the findings were analyzed and compared to current TM-related literature and promotional materials. Conclusions about the information in the corpus were drawn, recommendations for future research were made, and the methodology and overall project were assessed and discussed.

0.5 Outline

This thesis has been divided in to three chapters. **Chapter 1** provides a brief history of TM systems, and a critical summary of the literature related to TM functioning and usage. **Chapter 2** outlines the methodology used to select a type of primary source, identify specific primary sources, and collect and classify the data. **Chapter 3** discusses the contents of the corpus and compares this to the literature review in Chapter 1. Finally, the conclusion provides personal reflections, an evaluation of the methodology, and recommendations for future research. The thesis is also complemented by a list of references consulted, and appendices containing additional information regarding the structure of the thesis, the methodology, and the primary sources consulted.

Chapter 1 – The Translation Memory System

Chapter 1 is divided into seven subsections. Section 1.1 presents a brief historical summary of TMs. Section 1.2 provides a description of the components commonly included in TEnTs. Sections 1.3 and 1.4 summarize scholarly literature pertaining to considerations in purchasing and using a TM system, in order to examine the issues identified in an academic context. Section 1.5 describes an analysis of vendor promotional materials, in order to examine the issues identified in a commercial context as they influence the choice of systems. Section 1.6 summarizes recent TM-usage surveys, in order to examine issues in TM use and choice identified by researchers interacting with working translators. Finally, section 1.7 describes an analysis of translators' contributions to a mailing list, in order to provide some insight into their spontaneous feedback about TM tools.

1.1 Translation memories: Inception to adoption

1.1.1 The Birth of TM systems²

Technologies for translators can be categorized into two main types: machine translation (MT) and computer-aided translation (CAT). MT can be defined as “the process whereby a *computer program translates* a text from one natural language into another” (Bowker 2002, 3) [emphasis added]. In CAT, the *human translates* the text but calls upon the computer to assist with specific tasks.

After a period of intense MT research in the 1940s and 1950s, numerous factors contributed to the end of most large-scale MT projects (as identified in the ALPAC Report (ALPAC 1966)). Research in the area reached a virtual standstill until the 1980s, when researchers began to realize the potential of and investigate tools to assist in the translation process, rather than tools that perform the translation function themselves. In 1980, Kay wrote a report entitled “The Proper Place of Men and Machines in Language Translation” (later published as Kay (1997)), which encouraged a transfer of research focus from MT-based projects to other useful software applications for translators. His

² Only a very brief historical overview of the development of TMs and TEnTs is provided here. For a more detailed description, consult Hutchins (1998).

proposed tool, the *Translator's Amanuensis*, consisted of a text editor and a dictionary and was the first conceptualization of today's TM system (Kay 1997). In 1981, Melby suggested the use of bilingual concordancing as an aid to translators and developed the idea of a translator's workstation, or a collection of tools to aid in the translation process (Melby 1981). Brian Harris's concept of "bi-text", "a single text in two dimensions" (i.e. a paired source and target text), also advanced the progress of TM systems (Harris 1988, 8). The first research-based implementations of TM systems occurred in the early 1980s, and the first commercial systems appeared on the market in the early 1990s. Users were quick to adopt these products when they were released, and soon many commercial tools were available (Somers 2003, 33).

1.1.2 TM systems today

Today's market and the enthusiasm surrounding the use of translation technologies continue to grow.³ A 1999 survey conducted by the Canadian Translation Industry Sectoral Committee forecasted strong and continued growth in the translation industry (CTISC 1999, 6 & 28). Furthermore, the report recognized numerous trends related to translation technologies: "The automation of the environment in which translations are produced is expected to continue its rapid growth. The demand for aids for translation is expected to grow at a high rate in coming years. More specifically, translation memories will be used to greater extents as aids for translation" (Ibid., 51). These observations and predictions regarding strong growth both for the translation industry in general and translation technologies in particular were also echoed in a report published by Allied Business Intelligence in 2002 (ABI 2002).

An analysis of job ads (Bowker 2004) also revealed the increased adoption of translation technologies. Job advertisements were collected for a period of three years and then analyzed according to the skills and qualifications required by employers (Ibid., 961). The number of jobs requiring knowledge of specialized translation tools increased from 8 in 2000, to 17 in 2001, to 30 in 2002, virtually doubling each year (Ibid., 970). Bowker writes, "If this trend continues, it will not be long before a significant number of jobs do indeed require this type of knowledge" (Ibid.).

³ For instance, at least 20 competing TM tools are available on the market, including Déjà Vu, MemoQ, MetaTaxis, MultiTrans, OmegaT, SDL Trados, and Wordfast, to name just a few.

TM systems are becoming a more integral part of the profession, and the demand and need for their use will only grow (Rode 2000, 13). In short, TM systems are here to stay, and research in this area is more valuable than ever. Current trends in TM usage are further discussed in regards to the findings of recent TM-related surveys and investigations in sections 1.6 and 1.7.

1.2 Components of translation environments

As described in the *Scope and Limitations* (section 0.3), TM systems are not generally used in a stand-alone fashion, but rather as a part of a larger collection (or suite) of tools called a translation environment (TEnT). While the focus of this research is restricted to TM systems, it is necessary to understand and consider all of the components used with TM systems. The following section will provide definitions and explanations of the most common components.⁴

1.2.1 TM systems

TM systems are usually the main or central components of TEnTs. TMs are “databases that store translated texts together with the corresponding original texts” (Austermühl 2001, 135). A TM system searches these databases and allows a user to take advantage of previously translated material. When translating a new text with a TM system, the user moves through the text segment by segment, and at each segment, the system searches the TM database and proposes matches (i.e. exact or similar sentences that have already been translated) to the user. The translator then accepts, modifies, or rejects the match or matches proposed, or, if there is no match, translates the sentence from scratch. The TM is then updated (either during or after the translation process) to include the new source text and translation. The types of matches and other specifics of TM system usage are discussed in detail in section 1.4.

⁴ While localization tools are not *components* of TEnTs, it is important to note that many localization tools have some of the same components as TEnTs (e.g. TM systems, terminology management systems). The scope of this thesis does not exclude localization tools, but limits the focus of the investigation to the TM component of localization tools.

1.2.2 Other TEnT components

Table 1 provides brief descriptions of the most common TEnT components and how they are used with TM systems. The individual tools included in a TEnT and their robustness will differ from brand to brand. Refer to Bowker (2002), Quah (2006), or L’Homme (2008) for further information on TEnT components.

Component	Explanation
Bitext aligner	A bitext aligner performs a three-step process of file identification, segmentation, and alignment to create bitexts from source and target documents. Corresponding texts are matched up segment by segment. Bitexts are the basis of TMs, and aligners allow translators to create new TMs or add legacy translations to existing TMs.
Concordancer	A concordancer searches a text or bitext for a user-specified character string. Concordancers can be useful for finding and examining previous translation choices for words, structures, or terms (i.e. for finding equivalents that are shorter than the segments typically processed by TM systems).
Term extractor	A term extractor analyzes a (bi)text or group of (bi)texts and attempts to identify potential terms. Extraction may be monolingual or bilingual.
Terminology management system (TMS)	A TMS aids in the storage and retrieval of terms and terminological information. The contents form a term base. Term bases are often used during translation in conjunction with TM systems.
Active terminology recognition	An active terminology recognition tool scans a text in the source language, consults a specified term base, and automatically replaces any terms in the text with their target-language equivalents as contained in the term base.
Statistical document analysis	A statistical document analysis tool compares a new text to be translated with a given TM and/or term base. The results usually provide an overall word count and a separate count of segments found in the TM, and indicate the types of matches found in the TM (see section 1.4.3.1). These tools can be especially beneficial in deciding whether it would be useful to use a particular TM or term base when translating a text or for preparing estimates.
Project management module	Project management features can help users track client information, manage deadlines, and maintain project files for each translation job. TM maintenance, management, and quality control may also be addressed through project management modules.
Quality control module	Quality control features may include spelling, grammar, completeness, or terminological/controlled language compliance checkers.
Machine translation system	The inclusion of an MT system allows for the generation of a machine translation for a sentence to be translated when no match is found in the TM.

Table 1 - TEnT Components

1.3 Purchasing or acquiring a TM system

In the following subsections, I provide a critical summary of the current TM-related literature, in order to situate subsequent discussions of the corpus contents. Section 1.3 discusses the situational, technical, and brand-specific considerations in deciding to invest in a TEnT or TM system and in choosing a particular brand. While the following issues relate to TEnTs in their totality, the discussions are focused on the TM component of TEnTs. This subsection will focus specifically on the considerations that are most pertinent at the purchasing or acquisition stage. Other considerations related to the use of TM systems, which may include the type of file formats handled, availability and/or price of technical support, and compatibility with other TEnTs or TM systems, are discussed in detail in section 1.4. Vendor promotional materials and their role in tool selection are addressed through an analysis of vendor advertisements and websites in section 1.5. For the sake of clarity, issues are discussed in only one section; however, they may be relevant at any or all stages in the process of choosing, purchasing, and using a TEnT or TM system. The hierarchical structure of sections 1.3 and 1.4 is reproduced in Chapter 3 and the conclusion. Appendix A contains a table listing the issues addressed in these three places and cross-references to locations of discussion within the thesis.

1.3.1 Situational considerations

The discussions relating to situational considerations (i.e. to the type of work being done and the conditions of the work) focus on text characteristics and format, the number of users, language pairs, and language of the interface.

1.3.1.1 Text characteristics and format

The first decision potential TEnT users must make is whether the tool will benefit them in their work, an evaluation which may be based in part on text characteristics and format.

Pertinent **text characteristics** for TM system usage include the repetitive or non-repetitive nature of a text, the text's author or creator, the subject field or domain, the style of the text, and the text type. Existing literature states that TM systems are primarily used and are most beneficial when translating repetitive texts (Heyn 1998, 124;

Rode 2000, 12; Austermühl 2001, 139; Bowker 2002, 112; Quah 2006, 94). A TM system proposes previous translations of repeated segments (either from within the same document or from another document) for reuse, thereby reducing the amount of time needed to translate. In addition, other factors may contribute more or less directly to repetition; therefore, the use of a TM system may also be advantageous for translators who consistently work in a specialized field or for a particular client, or regularly update and revise translations. The benefits of using a TM system are directly related to these factors, and potential users may evaluate how the use of a TM system could reduce and improve their current translation work based on the text characteristics of the documents they translate.

In addition to content, client, and subject-field characteristics, the format of the texts to be translated must also be considered. In order to use a TM system, the translator must be working with electronic files. While electronic files may seem to be the norm in today's digital marketplace, hard-copy documents are still being used. The findings of two TM-related surveys show that the use of hard-copy documents was a commonly cited reason for non-adoption of TM systems (Lommel 2004, 6; Lagoudaki 2006, 20). If a potential TM user is working with hard-copy documents, he or she may need to convert them to electronic format, a process that obviously increases the amount of time needed to begin using a TM system. Two of the most common methods for conversion are optical character recognition (OCR) software, which recognizes patterns in scanned images and converts the images to text, and voice recognition software (VRS), which converts dictated text (speech) to written text in a text editor or word processor. While these methods may accelerate the conversion of hard-copy documents to electronic format, potential users should consider the time needed to complete this process (which may include additional time for proofreading and correcting errors that may be introduced during the conversion process) when evaluating the anticipated benefits and drawbacks of TM system usage. For further information on these forms of conversion, see Bowker (2002) and L'Homme (2008).

1.3.1.2 Number of users

The choice of a TM system or TEnT may depend on whether the system will be used by only one individual or by many users on a network. A TEnT may have no networking capabilities (i.e. the TEnT may be usable only by one user on one computer), or may support use on a network (i.e. the TEnT may be usable by multiple users, concurrently or separately) only in specific versions (i.e. corporate rather than freelance versions). A freelance translator will probably not need a TEnT with networking capabilities; however, agencies or organizations will most likely limit their investigation to those tools that can be used on a network. Potential buyers should be aware that various versions of a tool may exist and evaluate the possibilities accordingly, in order to choose the tool that best suits their needs and current working conditions.

1.3.1.3 Languages

1.3.1.3.1 Language pairs

The working languages of a translator must also be taken into consideration when investigating TEnTs. Unlike many MT systems, TM systems are, in principle, language independent in terms of their functioning. However, the use of a TEnT will only be beneficial if it supports the languages (e.g. character sets) in which the user is working (i.e. the source and target languages). Certain languages are processed more easily than others; for example, TM systems require further development for languages that read right-to-left (e.g. Hebrew or Arabic) or languages that use double-byte characters (e.g. many Asian languages).⁵ The choice of system based on language support may affect a freelance translator less than an agency or organization, as a freelance translator most likely translates in a limited number of languages, whereas the system used by an agency or organization will need to support all languages being processed by all of its translators.

Finally, a TM can be either unidirectional or bidirectional, and the system's directionality options may be a consideration; for example, a user may create a TM with

⁵ A byte is an electronic unit of storage. Basic Latin characters require only one byte of storage, whereas more complex characters, such as those used in many Asian languages, require two bytes of storage. These languages use a double-byte character set rather than a single-byte character set (Bowker 2002, 74). However, with the emergence of Unicode, a character-encoding standard, this problem may be less prevalent than in the past.

English as the source language and French as the target language but later wish to open the TM with French as the source language and English as the target language. A bidirectional system would be capable of processing the TM in either situation, whereas a new TM would have to be created for use with a unidirectional system. While many translators may work primarily in one direction, the reversal of the source and target languages could prove useful when working in the opposite language direction or conducting relevant research; although, as outlined by Bowker (2003), reversing a TM may have implications for both the quality and quantity of the matches that are retrieved.

1.3.1.3.2 Language of the interface

The language of the interface may also play a role when choosing a TEnT. An interface, or graphical user interface (GUI), is a “human-computer interface (i.e., a way for humans to interact with computers) that uses windows, icons and menus and which can be manipulated by a mouse (and often to a limited extent by a keyboard as well)” (LINFO 2004). In plain language, the interface is what the user sees and interacts with when using software. Although the language of the interface may seem relatively unimportant for bilingual or multilingual translators, “the comfort of being able to access tools’ options and documentation in one’s own language can make the experience of using a particular tool far more pleasant, and can shorten the learning curve” (Bowker, McBride, and Marshman 2008, 31). This may be especially important for users who are less comfortable with technologies or those who need to learn multiple tools (a common practice in the industry (Lagoudaki 2006, 23)). Users may be more comfortable operating in specific languages; therefore, the language(s) of a tool’s interface should be evaluated and considered, as this may directly affect the amount of time needed to learn how to use the tool and resolve technical difficulties.

1.3.2 Technical considerations

1.3.2.1 *Computing requirements*

1.3.2.1.1 *Processing speed and memory*

As outlined in section 1.2, TEnTs can be very large and multifaceted software applications. The memory requirements and processing specifications of a given TEnT should be noted, as the user must have an adequate computer system to use the application effectively, and additional upgrades to memory or processing speed will require further financial investment. Furthermore, the size of the files produced and accessed during usage should also be considered. TM files can be quite large, depending on the number of texts they contain, and perhaps especially in situations where users are working as part of a team and sharing TMs. A complex TEnT used with large TMs will require more memory and a more powerful processor than a less complex TEnT used with small TMs.

1.3.2.1.2 *Operating system*

An operating system (OS) is “the software that controls the execution of programs and that may provide services such as resource allocation, scheduling, input/output control, and data management” (TERMIUM 2006a). The OS is “the computer’s master control program” and interacts with, manages, and supports all files and software on a computer (Computer Language Company Inc. 2009). For this reason, it is important that a user installs only software that is compatible with or supported by the computer’s OS. The most common OSs are Microsoft Windows, Apple’s Macintosh OS X and Linux- or Unix-based systems.⁶ Software may be compatible with all, only one, or a combination of these OSs.

While Windows is the most widely used OS, the number of Macintosh and Linux users is growing (Smith 2005; Hesseldahl 2006), and as a result, more CAT tools are being developed for Macintosh and Linux-based systems. However, users of Macintosh and Linux OSs may have to investigate further to find cross-platform tools. Some brands are available in separate versions which are specifically designed to be used on Windows,

⁶ Linux is an open-source operating system. For a definition of open-source software, see section 1.3.2.2.

Linux, Mac, Solaris, or Unix OSs (see Heartsome 2008; OmegaT 2009b). As pointed out by Bowker, McBride, and Marshman, the availability of cross-platform tools and the option to install a TM system on the user's preferred OS may "lead to an increased uptake of the tools in translation work" (2008, 38).

1.3.2.2 Open-source versus closed-source software

The term **open-source software** refers to software whose source code can be accessed and modified by all users, allowing individuals or organizations to improve or customize the software. In addition, this software can be redistributed without any restrictions and is usually available free of charge. (See Open Source Initiative (2009) for the complete definition of open-source software.) The term **closed-source software** (also known as proprietary software) refers to software whose source code can only be accessed, edited, or redistributed by the owner and not by users. Closed-source software is primarily purchased; however, it may be available free of charge.

The increase in popularity of open-source software has reached the CAT tool market, and open-source CAT tools and TM systems (e.g. OmegaT) are now available options for translators. Many open-source applications are available for a range of OSs (rather than just one OS), and this availability may provide non-Windows users with additional options when choosing a TEnT.

1.3.2.3 Word-processor or text-editor integration

A word processor is an application that allows a user to create, read, edit, and format electronic documents. Examples of word processors are Microsoft Office Word, OpenOffice Writer, and Corel WordPerfect. A text editor can be broadly defined as "software that enables a user to create and revise text" (TERMIUM 2009b), and while text editors enable a user to modify text, they do not usually provide the same formatting or word-wrap features as a word processor. Some TM systems are integrated into a third-party word processor or text editor, and in these instances, the user interacts with the TM system through the interface of the word processor or text editor (i.e. use of the TM system *requires* the use of a third party word processor or text editor); therefore, in order to use a TM system that has this integration, the user must also own or obtain the word

processor or text editor. Bowker notes that the integration of a TM system within a word processor “can reduce the translator’s learning curve as he or she can continue working within a familiar piece of software” (2002, 123).

TM systems that are not integrated into a third-party word processor or text editor have their own text editor (a proprietary text editor). In these instances, the user translates directly within the TM system’s interface. This may negatively affect the user’s learning curve, as he or she will have to learn how to operate within the proprietary editor (L’Homme 2008, 181).

While the use of a word processor is perhaps unavoidable for translators, the issue of learning curve and word-processor integration relates to whether users will be required to learn a new word processor, work in a word processor other than their preferred tool, or learn how to use a proprietary text editor. Furthermore, using a system with an integrated word processor or text editor, as opposed to a proprietary text editor, may pose difficulties when accessing certain file formats. (Issues relating to file formats are discussed in section 1.4.2.)

1.3.2.4 Voice recognition software

Voice recognition software (VRS) may be useful not only for converting texts from hard-copy to electronic form as described in section 1.3.1.1, but also for producing texts and translations, and interacting with the computer (e.g. for executing commands such as opening, closing, or saving files). While most VRS packages are compatible with many of the most commonly used applications, those users who wish to use their current VRS with a TEnT should be certain that the applications are compatible.⁷

1.3.2.5 TM creation, storage, and usage methods

1.3.2.5.1 Sentence-based and character-string-within-a-bitext approaches

There are two main approaches to storing information in TMs. Macklovitch and Russell define the more narrow approach (sentence-based (SB)) as one that “fixes the sentence as the privileged processing unit of TM systems and presumes automatic look-

⁷ For an example of VRS, see the Dragon NaturallySpeaking website:
<http://www.nuance.com/naturallyspeaking/>.

up as the privileged processing mode” (2000, 137). The second approach (character-string-within-a-bitext (CSB)) is broader and “regards TM simply as an archive of past translations, structured in such way as to promote translation reuse” (Ibid.). The authors note that this definition makes no reference to how or at what level the information is stored or searched (Ibid.).

Each approach has its own advantages and disadvantages that must be considered not only when using the tool, but also when choosing a TM system. An advantage of the SB approach is that it operates based on easily identifiable units (sentences); the system can quickly and somewhat easily determine the beginning and end of each sentence according to formal criteria (e.g. capital letter at the beginning, punctuation at the end) (Gow 2003, 23). Moreover, the SB approach facilitates exchangeability of TMs among systems and users (Ibid., 24), as the majority of systems on the market use this approach (and not the CSB approach). Finally, the matches are generally of high precision (i.e. the matches retrieved are most likely relevant), as the match is calculated based on an entire segment (Ibid., 25). Difficulties include alignment problems caused by differing sentence structure between the source and target texts (Ibid.), the shift to translating segments or sentences rather than paragraphs and texts (Ibid., 26), and the limitations of matching algorithms (Ibid., 27). (See section 1.4.3 for additional discussion relating to the matches proposed by a TM system.)

The CSB approach allows the user to see a broader context of the match (i.e. the whole paragraph or text rather than just a single segment) (Gow 2003, 36). Furthermore, the user can take advantage of larger chunks of repeated or similar text (e.g. paragraphs, pages) rather than only operating on the sentence level (Ibid., 37). However, a TM system using this approach may produce more noise in the form of sub-segment chunks (Ibid., 38). Additionally, the translation of the current text is usually not added to the TM or verified until after the translation process is complete, thereby preventing the user from easily taking advantage of internal repetitions (Ibid.). Finally, TMs created using the CSB approach are less easily transferred among systems and users (Ibid.). Potential buyers may wish to investigate the benefits and drawbacks of these two methods and the consequences for their specific translation needs.

1.3.2.5.2 Interactive creation and post-translation alignment

Interactive TM creation is the process of adding to a TM as the user moves through the source text and translates the segments. When a target segment is inserted into the target text, it is automatically added to the TM and can be accessed immediately. Use of this method to create a TM may entail a significant time investment in order to build up a substantial TM, as translation units (TUs) (i.e. source- and target-segment pairs) are only added to the TM as the user produces translations. Furthermore, as noted by L'Homme, the user must either produce a satisfactory translation right away (as it is immediately added to the TM for reuse) or ensure that the TM is updated to include the corrected or edited segments (2008, 178).

Post-translation alignment allows translators to add legacy translations (i.e. previously completed translations) to the TM (L'Homme 2008, 179). The texts are segmented, and the corresponding source and target segments are automatically paired using a bitext aligner. While this method allows a user to create a TM quickly, the bitext produced by the aligner may need to be verified for correctness, as an incorrectly aligned bitext may produce noise (i.e. target segments that are not translations of the source segments) when used with a TM system. Furthermore, it is important to note that in the majority of TM systems using this method, the user is unable to benefit from internal repetition, as the target segment is not added to the TM until after the translation is completed. The relationship between interactive creation and post-translation alignment is not mutually exclusive; many TM systems (e.g. SDL Trados) provide both options for TM creation.

If a translator has a significant number of legacy translations and wishes to leverage the content of these translations, he or she may prefer using a system that allows for post-translation alignment (in addition to interactive creation) in order to reduce the amount of time needed to add texts and translations to the TM. Conversely, a translator who has fewer (or no) legacy translations may only make use of the interactive creation method, thus growing TMs as he or she completes translations.

1.3.2.5.3 Pretranslation and interactive translation modes

Pretranslation is a process during which the TM system scans the new source text in its entirety and automatically replaces any segments for which an exact match is found in the TM; segments for which no exact match is found remain in the source language. The user is not consulted during this process; he or she does not decide on an individual basis whether or not an exact match should be accepted. The result of the pretranslation process is a hybrid document which contains segments in both the source and target languages (L'Homme 2008, 180). For obvious reasons, the matches inserted during pretranslation must be examined carefully, and the translator often must refer to the original source text to ensure that the match is appropriate for the context. A translator's desire to use pretranslation may depend on his or her confidence in the quality of the TM. This confidence may be based on the accuracy of alignment, the source of the TM (personally created or received from another source), and/or its relevance to the text being translated. The use of exact matches requires that the translator evaluate whether the segment is appropriate in the current context, and this evaluation affects the amount of time needed to complete a translation. (This issue is discussed further in section 1.4.3.2 in relation to the usefulness of matches.)

In interactive translation, the translator moves through the source text segment by segment, and matches found in the TM system are proposed to the user. A match can be modified, accepted as-is, or rejected. Segments for which there is no match are translated from scratch. This method allows for immediate evaluation of matches, thereby decreasing the possibility of incorrect or out-of-context matches being inserted in the target text. (For a detailed discussion on the effects of these two methods on productivity, quality, and translator satisfaction, see Wallis (2006).)

All TM systems allow for interactive translation; however, if the user wishes to use a pretranslation function, he or she should investigate the availability of this feature in a system before purchasing it.

1.3.3 Brand-specific considerations

Numerous factors should be considered when choosing a particular brand of TM system, not only including investigations of the availability of trial or demo versions (section 1.3.3.1), the price (section 1.3.3.2), and update and new release considerations (section 1.3.3.3), but also the size, reputation, and history of the vendor. The latter three factors can be researched on vendor websites and in other literature (e.g. discussion forums, mailing lists) pertaining to the vendor. While these are all important considerations when choosing a brand, the following section focuses on trial or demo versions, price, and updates or new releases.

1.3.3.1 Trial or demo versions

As purchasing and using a TM system may involve a significant investment of time and money, it is important for potential users to investigate the available tools and decide which one best suits their needs. One of the best methods for this investigation is the use of a trial or demo version of the software, which would allow a potential buyer to evaluate the tool as it can be used in his or her particular translation context. Trial or demo versions are available for most TEnTs. They may be time restricted (i.e. usable for a certain period of time), size restricted (i.e. usable with a limited number of TUs in the TM) and/or feature restricted (i.e. usable with a limited number of features).

1.3.3.2 Price

As outlined in section 1.2, TEnTs can be extremely complicated and multifaceted software packages, and the price tag can reflect this. Price can depend on the components of the tool, its sophistication, its features, and/or its networking capabilities, and may range from free (e.g. open-source software) to over a thousand dollars, with numerous options in between.

Some tools are available in more than one version, such as a freelance version, a professional version, or a workgroup version, and the price varies accordingly (see discussion of number of users in section 1.3.1.2). Furthermore, some vendors offer free versions of their tool, which often have limited features, offer read-only access to files, or only allow processing of projects (i.e. sets of files) created with the full version (enabling

clients/agencies to supply the projects to the translator for use on a free version of the tool).

While price must be considered when investigating TEnTs, it is not always indicative of a tool's potential to meet the needs of a particular user. Furthermore, while the expense of purchasing a TEnT can be high, the amortization of the investment will vary depending on the translators' type and volume of work, learning curve, and/or comfort with the tool. (See section 1.6.3.3 for a discussion of price and return on investment as inhibitors to TM adoption.) While the bottom line may be a deciding factor, all of the issues outlined in sections 1.3 and 1.4 should also be considered when choosing a tool.

1.3.3.3 Updates or new releases

It is important to keep a TM system up-to-date by installing updates or new version releases. In addition to adding new features, new versions of a tool may resolve bugs discovered in previous releases, and/or provide compatibility with new file formats or new releases of related software (i.e. word processors, text editors, or OSs); for example, when a new version of an integrated word processor or text editor (i.e. a word processor or text editor that is *required* for use of the TM system (see section 1.3.2.3)) or OS (see section 1.3.2.1.2) is released, a service pack or update may be provided by the TM system vendor. This update would eliminate the compatibility issues between the TM system and the newer version of the other software. While new versions may resolve bugs in previous versions and offer new or more advanced features, users may have to spend additional time on training in order to learn how to use the new version or new features.

The number of and time between version releases varies from tool to tool and may range from once every few months to once every few years. The requirement to purchase any or all new releases may lead to instances of vendor lock-in (see section 1.4.2.3), as users may be unable to process newer versions of the tool's proprietary formats (see section 1.4.2.1.3) or receive technical support for older versions of the software. Depending on the vendor's policy, new releases or updates of TM systems may be available free of charge indefinitely or for a limited period of time (e.g. two years), or

they may only be available when purchased (usually at a lower price than a new licence of the tool), and users should take any additional costs into consideration when evaluating the initial price of a tool.

1.4 Using a TM system

This section describes the following issues related to using a TM system: training, file formats, matches proposed by a TM system, procedural and textual effects of TM usage, and client relations. These issues do not only relate to the use of a TM system or TEnT, but also may play a role in choosing a tool.

1.4.1 Training

After choosing a TM system, the user must learn how to use the tool. As O'Brien points out, proper training is absolutely necessary in order to overcome a user's potential fear of the tool and to ensure that the tool is being used as efficiently as possible (1998, 121). The following section addresses issues of training based on time, availability, and price. (The issue of training is related to technical support, which is discussed in section 1.4.4.2.)

1.4.1.1 Time

As outlined in section 1.2, TM systems can be highly complex computer applications, and, as a result, an investment of time is required to learn how to use them properly and efficiently. The learning curve is directly affected by the user's comfort and familiarity with computers and computer applications, and any previous experience he or she may have using TM systems. Somers writes that initially "you will need to invest considerable effort in setting up your TM system and learning how to get the best out of it" (2003, 43). During this (potentially demanding) learning phase, the translator can expect to produce less, as more time is required to operate and become comfortable with the TM system. Furthermore, a translator may use more than one brand of TM system, potentially lengthening and complicating the process, especially if he or she is learning how to use the tools simultaneously.

1.4.1.2 Availability

While most TM systems are accompanied by manuals and help files, this type of information may not provide sufficient training materials for all users. A user who is not comfortable with computers or software may find it difficult to simply read the manual in order to learn how to use a tool. Furthermore, TM systems function by “recycling” information from databases and other resources that are typically generated by the user, and it may be difficult to learn how to use a system (and ultimately generate these) without first having access to some resources. As a result, some vendors provide sample files which can be accessed and used for training purposes and/or in the context of tutorials.

Receiving adequate training may be less of an issue for an in-house translator than a freelance translator, as the employer may provide training sessions, whereas a freelance translator must find training him- or herself. TM vendors may provide step-by-step tutorials, webinars, and training guides which can visually aid the user in completing the necessary tasks. Formal training courses for groups or individuals are also offered by some vendors. Courses may focus on beginner or advanced users, and the length and intensity of a course depends on the vendor and the type of session chosen. In addition, professional associations may also provide workshops devoted to TM tools training. While these types of courses are perhaps more effective methods for training, as they involve human interaction, they are usually not available free of charge. In recent years, translator training programs have begun to incorporate translation technologies into their curriculum. This development does not affect those who have already graduated; however, it may impact TM awareness and knowledge among future generations of translators. While translator training programs cannot provide training on all brands of TM systems, students who have received training during a translation program may be more comfortable with the principles of TM tools. A TM user should consider the format of training that best suits his or her needs, investigate where/how he or she may be able to receive this training, and evaluate the possible time and financial investment it would entail.

1.4.1.3 Price

Training can be quite expensive, especially if the user requires individual sessions or training on focused topics. Group training can be more economical; however, the trainer may not be able to address each user's specific needs. Prices will vary depending on the number of attendees, the level of the session (e.g. beginner or advanced), and the length.

While training can be expensive, a good working knowledge of a tool can prevent or reduce the number and/or severity of technical difficulties encountered later, and an up-front investment in training may benefit the user in the long term. The price of training should also be taken into consideration when evaluating the initial price of investing in a tool.

1.4.2 File formats

TM systems both process and generate a number of different types of files, and each system reads and creates different formats and combinations of formats. Firstly, the format of source-text files may differ from job to job or client to client. These files may need to be converted using a filter in order to be accessed by the TM system. Secondly, a user may wish to align previous texts and their translations to create or add to a TM, and these files may be stored in a variety of formats. In addition to input file formats, the output formats should also be considered. Systems create proprietary files (e.g. TM files) during or after the translation process (aside from the finished translation). The file format of the final target text will most likely need to be the same as the source text, and conversion may require additional steps. Finally, specific systems are able to import and export TMs in one or many different formats. File formats are generally not universal, and each brand of TM system may have its own proprietary format. A lack of support for a diverse range of file formats may lead to vendor lock-in (section 1.4.2.3) and will definitely cause compatibility issues (section 1.4.2.2) when moving from system to system. The use of open standards (section 1.4.2.4) is one method of combating problems related to compatibility. The following sections discuss these issues (i.e. input and output files, compatibility, vendor lock-in, and open standards) in more detail.

1.4.2.1 Input and output files

1.4.2.1.1 Source and target file formats

It is important for current and potential TM users to consider the formats of source and target files with which they frequently work, as each new source-text file to be translated must be opened with the TM system, and target files will most likely need to be provided in specific formats. A TM system must be specifically designed to read and/or write a file format, or be equipped with a filter to convert the original source file and then later reconvert the translation produced. Common file formats include those created by Microsoft Office programs (e.g. .doc, .ppt, .xls), .pdf files created by Adobe Acrobat, tagged files written in hypertext markup language (HTML) and extensible markup language (XML), or system/executable files that allow programs to be launched and used (e.g. .exe, .ini). TM users should be aware of the TM system's file format capabilities in order to ensure that their file format requirements can be met by their current system.

In some files (e.g. .html, .xml, .exe, .ini), the text to be translated may be written between or within fragments of programming code. In these instances, most of the code does not get modified, and only selected elements of the file are translated. To increase the speed and ease of translation, and invoice/estimate creation for these types of files, a translator may wish to use a tool or TM system that is capable of filtering out the tags and presenting only the translatable elements.

Although only certain file formats may be naturally accepted or produced by TM systems, add-ins or filters can convert the files to the necessary formats. These filters must be specifically designed for the conversion of one format to another (e.g. .pdf to .doc). Bowker writes, "Ideally, a filter will extract the translatable text from the original file and present it to the translator in a translation-friendly format [...] [a] good set of filters will attempt to ensure that neither the content nor the formatting information is lost" (2002, 119). However, this is not always the case, and users must ensure that the conversion does not result in data loss or incorrect formats and formatting. Over the past few years, developers have been creating and improving filters, and this is one area of TM system development that will continue to be addressed in the future (Ibid., 138).

Additional steps required for conversion, and the availability and effectiveness of filters should be considered by each user based on his or her individual needs.

1.4.2.1.2 Legacy file formats

As described in section 1.2.2, bitext aligners are common components of TEnTs and can be used to align legacy source texts and translations. While these files must first be in electronic format, they must also be in formats that are accessible by the aligner; for example, an aligner may be unable to align texts in .pdf format, whereas .doc files may be processed by the aligner with no difficulty. An ideal situation would involve the alignment of texts without any conversion, as conversion may be time-consuming and require an additional step, or data may be lost or distorted during the process. Depending on the aligner, the files may need to be converted using separate software or filters before the alignment process can be started.

1.4.2.1.3 Files created by the TM system

Some TM systems convert files to proprietary formats (a format used with a particular system) for use with the TM system during or after the translation process. The proprietary files often include tags containing additional information about the text and how it was segmented or otherwise processed. For example, Translator's Workbench, the TM system in SDL Trados, creates hybrid files during the translation process. These hybrid files contain the target (translated) segments, accompanied by the matching, hidden source segments, distinguished by tags. After the translation has been completed, the user must perform a clean-up to create clean files that contain only the target segments (SDL 2009).

Uncleaned files, however, are not without purpose. Because the uncleaned files contain all matching segments as well as additional textual information, they can be used during the clean-up process to update, create, or add to TMs. Some clients may require translators to deliver both clean and uncleaned files, as using the uncleaned files to update a TM is more efficient and accurate than aligning source and target texts. In these instances, there is no need for a translator to supply a separate TM, as the uncleaned files contain the same information. For this to work, however, the translator must be

producing files that are compatible with the client's system. Issues of compatibility are discussed in section 1.4.2.2.

1.4.2.1.4 TM formats

The formats of TM files must also be considered. If a user is importing a previously created TM, the system must be able to access and read that particular format. Export capabilities are also important, especially in situations where a client requires the TM to be delivered in a particular format. The issue of importing and exporting TMs is also closely related to compatibility, and is discussed further in the following section.

1.4.2.2 Compatibility among TM systems

Compatibility among TM systems can be extremely important for translators who use numerous applications or for translators who use systems other than those used by their clients. While the use of filters, converters, and open standards (see sections 1.4.2.1.1 and 1.4.2.4) may provide sufficient means for this compatibility, in some instances these solutions are not available or adequate.

Some TM systems have been developed to handle and produce the file formats of other TM systems; for example, OmegaT can be equipped with a utility that converts SDL Trados .ttx files into a format that is compatible with OmegaT, enabling the user to translate these files in OmegaT without using SDL Trados (OmegaT 2009a). This site also describes a converter that allows OmegaT users to open TMs generated by Wordfast (Ibid.). Déjà Vu can also handle file formats created by SDL Trados, from both Translator's Workbench and TagEditor (ATRIL 2009).

Depending on the level of compatibility, a translator may be able to use the TM system of his or her choice for projects from clients using different systems. Furthermore, compatibility may allow users to import existing TMs and files into a new or different system, enabling access to legacy translations and previously created TMs. While compatibility increases opportunities to move among different TM systems, it is important to realize that the conversion and generation of these formats may result in the loss or incorrect processing of data, and 100% compatibility is rarely achieved.

1.4.2.3 Vendor lock-in

Vendor lock-in can be defined as “the situation in which customers are dependent on a single manufacturer or supplier for some product [...] and cannot move to another vendor without substantial costs and/or inconvenience” (LINFO 2006b). In the computer industry, this term is often used to describe situations in which there is no compatibility among file formats or OSs (Ibid.). ATRIL, the developer of Déjà Vu, has created a compatibility guide for using and producing SDL Trados file formats with Déjà Vu. The introduction to this guide advises that “when choosing a Translation Memory tool, [you should] make sure your vendor won't lock you in: make sure your software can open and use other vendors' files, but also that your hard-worked [*sic*] translation resources remain in standard, non-proprietary formats for future use by you or by others” (ATRIL 1999, 2). The freedom to transport files from one system to another may be especially important for translators working with multiple TM tools and negotiating brand requirements with numerous clients. While the use of open standards can provide users with this freedom, instances of vendor lock-in may persist due to the prevalence of a particular brand among clients, agencies, and/or translators, or pricing policies for updates and new releases of TM systems (see section 1.3.3.3).

1.4.2.4 Open standard formats

A standard can be defined as “a uniform set of specifications for some or all aspects of a product (i.e., [*sic*] a good or service) or activity that allows interoperation of the product with some other products (including inputs and/or outputs)” (LINFO 2006a). Open standards are freely available to the public and are not proprietary or vendor-specific. This availability encourages the development of software that can read, process, and produce data in the open standard format, and consequently, an open standard format is potentially readable and producible by *any* TEnT. There are currently numerous open standards in the globalization, internationalization, localization, and translation (GILT)

industries; however, due to the scope of this project, only those directly related to TM systems – TMX and SRX – are discussed in detail.⁸

Translation Memory eXchange (TMX) is an open standard that defines the format of TM databases to increase their transferability among different TM systems. TMX was developed in 1998 by OSCAR (Open Standards for Container/Content Allowing Re-use), a division of the Localization Industry Standards Association (LISA) (LISA 2008b). TM systems can be programmed to read and create TMs in TMX format, making this format a common data exchange standard for a variety of tools; for example, a TM created with Wordfast could be exported to TMX and then later used with Déjà Vu, or vice versa. The majority of TM systems currently support (i.e. import and export) files in the TMX standard.

The Segmentation Rules eXchange (SRX) format, another open standard developed by LISA, facilitates the transfer of information relating to segmentation rules or criteria (i.e. how the software decides where one segment ends and the next begins). SRX files are often used in conjunction with TMX files, allowing TM systems to “adjust their processing to emulate that of the tool that created a TMX file, thus improving leverage” (i.e. increasing the number of matches retrieved) (LISA 2008a). Users who require or anticipate the need for exchangeability of TM resources among systems may consider open standards as a method for this compatibility and wish to use tools that support open standards.

1.4.3 Matches proposed by a TM system

A TM system functions based on segment matches found in a TM and the new source text to be translated. These matches are identified through a process of character matching (i.e. a calculation of the number of character strings that are identical and in the same order in two segments or sub-segments). This calculation is expressed in the form of a *match value* (usually a percentage); for example, a pair of segments in which six of eight character strings are identical might have a match value of 75%. While the basis

⁸ Two other commonly used standards in the GILT industries are the Term Base eXchange (TBX) and the XML Localisation Interchange File Format (XLIFF) standards. These two *formats* may be used in conjunction with TM systems and TEnTs; however, they are not specifically related to TM systems.

for match identification remains the same from system to system, each TM system has its own algorithm for evaluating matches.

1.4.3.1 Types of matches

Table 2 contains explanations of the different types of matches that may be proposed to the user by a TM system. While not all TM systems provide each type of match described in the table, it is important for TM users to understand the types of matches, and the subsequent impact on time spent editing, revising, and translating, in order to choose a tool which best suits their needs and allows them maximize the usefulness of their previous translation work.⁹ For further explanation and examples, see Austermühl (2001), Bowker (2002), Quah (2006), or L'Homme (2008).

Type of Match	Explanation and Example
Exact	An exact match occurs when the new source segment to be translated is exactly (100%) the same as a segment in the TM. Example: <u>New source segment:</u> The channel shall be trapezoidal and have a bottom width of 1.0 metres. <u>Exact match retrieved from TM:</u> The channel shall be trapezoidal and have a bottom width of 1.0 metres.
Full	A full match occurs when differences are found only in specific types of information, such as numbers, dates, proper names, currencies, or measurements (Austermühl 2001, 137). These items may be referred to as placeables, because they are often either invariable or translated in regular, predictable ways and thus can relatively easily be inserted (placed) semi-automatically in the translated segment. Some systems can recognize placeables and do not consider differences in these items in the calculation of the match value. Example: <u>New source segment:</u> The channel shall be trapezoidal and have a bottom width of 1.5 metres. <u>Full match retrieved from TM:</u> The channel shall be trapezoidal and have a bottom width of 1.0 metres.

⁹ The examples in Table 2 provide only the source segment or sub-segment; however, within a TM system, these source-text segments or sub-segments from the TM would be paired and presented with the target equivalents from the TM.

Fuzzy	A fuzzy match is a partial match: it is similar to, but not exactly the same as, the segment in the new source text. A fuzzy match can have a match value ranging from 1% to 99%. In most TM systems, the user can choose his or her own fuzzy match value, or fuzzy threshold. Generally this value is set between 60% and 70% (Bowker 2002, 100); however, in some instances a match as low as 40% can be useful (O'Brien 1998, 117). Many TM systems highlight the areas of difference in a fuzzy match. Example: <u>New source segment:</u> The channel shall be elliptical and have a bottom width of 3.28 feet. <u>Fuzzy match retrieved from TM:</u> The channel shall be trapezoidal and have a bottom width of 1.0 metres.
Sub-segment	A sub-segment match is based on a consecutive chunk of text that is shorter than a complete segment, although it is often longer than a single term. These sub-segments do not necessarily correspond to semantic units. While exact and fuzzy matches are calculated in relation to the whole segment, sub-segment matches are calculated based on small chunks of a segment. Not all TM systems identify sub-segment matches, as the process is more complex than segment matching and requires additional programming. Example: <u>New source segment:</u> The channel shall be trapezoidal and have sideslopes of 2H:1V that shall be covered with topsoil and seeded with an approved mixture. <u>Sub-segment matches retrieved from TM:</u> channel shall be trapezoidal seeded with an approved mixture
Term	If a TM is used in conjunction with a term base, the system may propose term matches found in the term base. These matches, as well as sub-segment matches, may be useful as a starting point when no exact, full, or fuzzy matches are identified.

Table 2 - Types of Matches Identified by a TM System

Fuzzy, full, and even exact matches require careful inspection and/or revision, perhaps with even more attention to detail than a translation from scratch. As de Vries points out, "Matches in [one language] are by no means always [...] matches in other languages!" (2002, 46). In cases where TM matches are used, careful post-editing is required to ensure that the translation is grammatically correct, idiomatic, and appropriate for the context. See section 1.4.5.2 for further discussion of the textual effects of TM usage.

Of course, not every segment will have a match in a TM. In these cases, the translator usually translates the sentence from scratch; however, in recent years, MT systems have been integrated with TM systems, thus providing the user with an automatic translation of the source segment, which he or she can accept, edit, or reject. Even more recently, some TM systems have begun to introduce example-based machine translation (EBMT), an approach to generating automatic translations by examining bitexts, or previous source and target texts. In EBMT in TM systems, the system would attempt to generate a target segment based on fuzzy, sub-segment, and term matches. The use of automatically generated translations (i.e. from an MT or EBMT system) may further reduce the time spent translating, although it might increase the amount of post-editing required. While some translators may prefer to edit and revise proposed translations, others may rather translate from scratch; based on this preference, users may choose to work with certain types of matches more or less frequently.

1.4.3.2 Usefulness of matches

The type, number, and quality of matches depend not only on the size and contents of the TM, but also the organization of TMs (i.e. the creation of specific TMs for individual clients or particular subject fields). Newer users may be more likely to develop one large TM containing all of their texts and translations; however, this approach may not be the most beneficial (Bowker 2002, 108). The division of TMs based on subject field and/or client may result in the retrieval of fewer matches; however, the matches will most likely be more relevant to the context. Preferences on how to organize TMs (i.e. one large TM versus many smaller and more specific TMs) are entirely personal, and each individual may see different benefits or drawbacks to the two approaches. Gow notes that less experienced translators rely on the contents of TMs not to save time, but rather to provide translation solutions, whereas “an experienced translator may be more interested in the approach that requires the least amount of time” (2003, 104).

The usefulness of matches proposed may also be related to the origin of the TM (i.e. personally created or received from an agency, client, or colleague). The content of TMs received from agencies, clients, or colleagues may not be beneficial due to

variations in subject field, client preferences, or style, and reviewing and editing these potential matches may require additional time. (For further discussion see section 1.4.6.3 related to ownership and section 1.4.6.4 related to buying/selling/trading TMs.)

Finally, the usefulness of fuzzy matches is usually directly related to the fuzzy match value, which is most often chosen by the user. Again, personal preference and the specifics of the source text and TM will dictate where a user may set the fuzzy match threshold. There is no standard for the fuzzy match value; however, a significant amount of research has been conducted in this area. (See Fifer (2007) for a detailed discussion of fuzzy match values.)

1.4.4 Technical issues

1.4.4.1 Customizability of TM systems

Depending on the type and brand of TM system, a user may have many, few, or even no options for customization; for example, the programming code of open-source software can be adjusted to meet a user's or group of users' specific needs, whereas closed-source software offers no source-code customizability. While the use of open-source software may allow the greatest amount of flexibility, the modification of source code is extremely technical and generally only performed by those with computer programming experience or knowledge. As such, discussions of customization focus not on programming alterations but rather on the customization options affecting the display and keyboard shortcuts within a TM system.

A user may wish to modify the default display and keyboard shortcuts in a TM system in order to increase the ergonomics of the GUI, make information more visible or more easily accessible based on personal preferences, facilitate the creation of a more streamlined process through enhancements to the user-friendliness of a tool, or align a TM system's display and shortcuts with those of another system the user has previously used and to which he or she is accustomed.

Display customization may include the rearrangement of windows or information within windows, modifications to font colours, and the addition or removal of options or icons to/from toolbars and/or menus. Keyboard shortcuts (sometimes called hot keys or accelerator keys) provide an alternative method of executing commands and interacting

with the interface, rather than using the mouse. As a user becomes more comfortable with a TM system, he or she may notice repeated use of certain commands and wish to increase the speed with which these commands are executed by adjusting current keyboard shortcuts or creating new ones.

Not all TM systems allow for modifications to keyboard shortcuts or the display; therefore, the importance of the level of customization should be established prior to choosing a tool. While customization may be somewhat less important to a new TM user, as he or she becomes familiar with the system or transfers from a known system to a new one, customizable options may prove more valuable.

1.4.4.2 Technical support

Technical support may be provided by tool vendors or other users of a TM system. Not only the provider, but also the form, availability, and price of technical support vary from brand to brand.

Vendor-provided technical support may be available through e-mail, over the phone, or online, and a user may be required to purchase a technical support contract in order to receive support from the vendor. Technical support contracts may be for any period of time, but most likely must be renewed annually or biennially. Contracts will also vary in price, depending on the type of support provided and the length of the contract. Numerous vendors (e.g. SDL Trados, Wordfast, MultiTrans) provide and maintain free resources such as knowledgebases or collections of frequently asked questions (FAQs). These databases contain previously identified issues and their resolutions and can be accessed freely by any user. Support may also be found in the form of user groups (e.g. Yahoo!, Google), discussion forums (e.g. ProZ.com, TranslatorsCafé.com), or mailing lists (e.g. Lantra-L, brand-specific lists). These forms of support are provided by other users of a particular TM system who voluntarily help their peers.

While the options for free support may be more appealing, users should consider their preferred format for support (e.g. speaking with an individual over the phone, receiving written instructions via e-mail), the quality of the support provided (e.g. by the vendor or developer versus by a fellow user of the tool), and the response time (e.g.

phone or e-mail availability of a technical support representative or online availability of fellow users). Before purchasing a TM system, users – particularly beginners or those less familiar with computers – should seriously investigate technical support options. In addition, it may be valuable to research the reputation of particular developers or vendors for giving prompt and helpful support.

1.4.5 Procedural and textual effects of TM usage

1.4.5.1 Speed

Speed of translation is one of the most commonly cited benefits of TM usage (Austermühl 2001, 135; Bowker 2002, 115; L’Homme 2008, 173). When translators do not have to translate from scratch and can benefit from previously translated work, they may see an increase in the speed with which they translate, as well as a decrease in the amount of research they must conduct. For clients, this may lower costs, decrease the time-to-market for translated or localized products, or shorten the process. While the idea of higher income and productivity is tempting, users should be aware that they may see an initial drop in productivity as they become familiar with the tool and build and develop their resources or TMs (Bowker 2002, 115). A TM system does not come equipped with TM databases; TMs must be created and added to by the user. Furthermore, a small and relatively ‘young’ TM may not prove as useful as a TM that has been enlarged over a long period of time. For these reasons, a TM user creating his or her own TMs “from scratch” is more likely to see the return on investment (ROI) not in the short term, but rather in the middle or long term (Ibid., 116). This, however, may not be the case if a translator is working with a TM which has been provided by a client/agency, purchased (see section 1.4.6.4), or previously created with another TM system.

1.4.5.2 Quality

Translators who use TM systems may translate repetitions more uniformly, allowing for greater consistency between versions of a text and among different texts created by the same client (O’Brien 1998, 119). However, the information in TMs must

be up-to-date, correct, accurate, and adjusted to fit the context in order to produce high-quality and consistent translations (Austermühl 2001, 140).

It has been observed that the use of TMs can negatively affect the quality of translations by potentially interfering with the internal cohesion of a text and reducing it to a collection of segments, sometimes referred to as **peephole** translations (Heyn 1998, 135). In the article “Orwellian Linguistics”, Mogensen writes that technical translators are “most often producing texts with a syntax close to that of the source text” in order to ensure high match values and that “language is now being changed to fit the tools, instead of the other way around” (2000, 29). The resulting structure affects the style of the translation and may detract from the naturalness of the text produced. Furthermore, as TMs operate on the segment level, some translators believe that the idea of a *text* no longer exists. In order to regain the importance of the text, some TM users have “switched from sentence-based to paragraph-based segmentation” (Bowker 2002, 118). As discussed in section 1.4.3 relating to matches and TM systems, the cohesion of the text must be addressed when dealing with any type of match from a TM system, a step which may be particularly important depending on the division of TM content and/or the source of the TM (see section 1.4.3.2). Refer to Bowker (2006) for further discussion of textual effects of TM usage.

1.4.6 Client relations

1.4.6.1 Brand or format requirements

As mentioned in section 1.4.2.1, translators working with clients who use TMs may have to consider the brand or file format requirements of the client. A client may provide the translator with a TM for a certain project, in which case the translator must be able to open, use, and potentially add to the TM. This may not be problematic if both parties are using the same brand of system; however, if the client and the translator use different brands of TM systems, the translator may have to convert the TM database in order to access it and/or modify its contents. Furthermore, the client may require the translator to return not only the finished document, but also uncleaned, tagged, or proprietary files created during the translation process. These requirements may influence the brand of system a translator chooses to use, or create compatibility issues

between the translator's chosen brand and the client's brand, an issue which may be further complicated if the various clients of a translator require different formats and brands of systems. For a complete discussion of issues related to file formats and compatibility, see section 1.4.2.

1.4.6.2 Payment

Payment practices have been affected by the adoption of TM systems. Statistical document analysis tools (see section 1.2.2) can analyze a new text to be translated and provide users with the number and types of matches found in a TM. Based on these results, clients may request or impose lower rates of pay for the matches found. In fact, some clients may be unwilling to pay the translator at all for any exact matches found in the TM. Austermühl notes that “translators are paid less for the ‘translation’ of fuzzy and exact matches”, potentially charging 30% of the standard rate for 95-99% matches, 60% of the standard rate for 85-94% matches, and the full rate for matches with a value of less than 85% (2001, 140-1).

Literature addressing the issue of payment reveals that translators may be dissatisfied with this arrangement and feel that they should be paid their full rate for all matches, as the use of exact, full, and/or fuzzy matches does not necessarily reduce the amount of time needed to complete a translation (see discussions in sections 1.4.3.2 and 1.4.5.2). Some translators feel that they should be additionally compensated as the use of TMs can improve the overall quality and consistency of their translations (Ibid., 122). In the article “Getting Full or Fuzzy?”, de Vries writes that “discounts have grown and were extended to even lower categories of fuzzy matches, meaning that discounts now often apply to sentences which need so much rewriting that they are practically translated from scratch” (2002, 45). As a result of these payment trends, some translators may choose not to use TM systems or decide to charge by the hour rather than by the word in order to compensate for the additional work involved in using a TM (e.g. file processing, conversion, maintenance) (Bowker 2002, 12).

1.4.6.3 Ownership

TMs can be valuable sources of information, and, as such, “both translators and clients are naturally anxious to claim ownership” (Bowker 2002, 122). Questions of copyright may arise when ownership of a TM is not clearly defined; while both translator and client have legitimate claims, there is currently no standard agreement on the rights to a TM.¹⁰ A client who owns a TM will most likely pass the TM along to other translators; Bowker notes that in these cases, it is possible for translators to “have the benefits of working with a TM even if they did not create it themselves” (Ibid., 123). However, this arrangement may not be ideal, as the translator may “find it constraining to have to adopt the style and wording of material that has previously been translated by someone else” (Ibid.). While clients may redistribute their TMs among their translators, translators must consider the ethical implications of using a TM supplied by one client to translate texts for another client.

1.4.6.4 Buying/selling/trading TMs

Closely related to the issue of ownership is the concept of buying, selling, or trading TMs, allowing TM users to benefit from the TM resources of other translators, agencies, or organizations. Again, issues of copyright come into play here: Can translators sell TMs which contain both the client’s original source texts and their translations? Can clients sell TMs containing translations which were not done by them? Obviously these legalities and issues of confidentiality complicate the development of a TM marketplace. Furthermore, the applicability of an acquired TM must also be taken into account; for example, would an organization in the health sciences truly benefit from a TM created by an organization or company in the physical sciences? (Also see section 1.4.3.2 on the usefulness of matches.) In addition, the quality of the contents of a TM can also be an obstacle when buying, selling, and trading. In an investigation of the possibility of TM exchange, Topping notes that when asked about the effect on style, “some translators respond that they would use the additional databases as reference materials, and that translation databases would not be applied without ensuring that quality was good” (2000, n.p.).

¹⁰ For a more detailed discussion on this complex legal issue, refer to Gow (2007).

The findings from two surveys (Lommel 2002, 16; Lommel 2004, 16) revealed little interest in the idea of buying, selling, and/or trading TMs due to the legal, confidentiality, and applicability concerns mentioned above. (For a discussion of these surveys see section 1.6.)

In spite of these roadblocks, one company, the TM Marketplace, sells and purchases licences for TMs. These TMs contain short source and target segments (approximately 10 words long). The company's *White Paper* states that "since the distributed database is comprised of short phrases of no more than 10 words of factual or technical information, each such phrase is not copyrightable because it is insufficiently original and protectable under the copyright laws" (TM Marketplace 2009, 1).

The Translation Automation User Society (TAUS) Data Association (TDA) is another example of an organization that has overcome the issues involved in TM exchangeability. The TDA is "a global not-for-profit organization providing a neutral and secure platform for sharing language data" whose members share TMs and term bases; each member provides their databases and in return can access other members' databases (TDA 2009a). The TDA, established in mid-2008, currently has over 40 member organizations (TDA 2009b).

These examples demonstrate that the buying, selling, and trading of TMs appears to be becoming more popular as organizations find ways to address the surrounding issues.¹¹ Interest in this area may continue to grow and develop considerably in the near future.

The previous sections have outlined the issues relating to choosing a TM system, purchasing or acquiring a specific brand, and using a TM system. These issues have been discussed and highlighted with examples and quotations from current literature. This, however, is not the only method for investigating these issues. The following section provides a brief analysis of vendor promotional materials (i.e. websites and advertisements), in order to reveal complementary information and allow for comparisons.

¹¹ Another example of TM exchangeability is the European Commission's 22-language TM of the *Acquis Communautaire*, which was made publicly and freely available in 2007 (EC JRC 2008).

1.5 Vendor promotional materials

In order to identify the promotional points highlighted by TM system vendors, an analysis of TM system advertisements and website literature was performed. Advertisements and websites are designed to inform and convince the reader to buy a particular product; therefore, an analysis of their contents will reveal the topics and issues that vendors believe to be the most important or necessary in the decision-making process.

This analysis included a total of 33 sources: 12 advertisements from the language technology magazine *MultiLingual* and 21 websites of TM system vendors.¹² The information contained in these advertisements and on the websites was classified according to topic. Table 3 lists the most common topics accompanied by the number of advertisements and/or websites that mentioned the topic. See Appendix B for the complete reference information.

¹² Five issues of *MultiLingual* were examined (see Appendix B for a list of the sources consulted in this section) and one advertisement per TM system was chosen. The websites of all TM systems mentioned in the corpus of discussion postings (see Chapter 3) were also investigated. Vendor websites were not investigated in their entirety; only the homepage, or, when available, the main page of the TM system component were analyzed.

Topic	# of References in Advertisements and on Websites (33 sources total)
File Formats Handled	19
Networking Capabilities	14
Interface of Tools	14
Compatibility with Other Tools	14
Increased Speed of Translation	13
Training	11
Open Standards	11
Localization Tasks	11
Improved Translation Quality	11
Variety of Versions	10
Technical Support	9
Price	9
Operating System	9
Clients of Vendor	7
Lower Translation Costs	7
Matching Capabilities	7
Customizability	7
Languages Handled	6
Other Services Offered by the Vendor	5
Increased Income	5

Table 3 - Top 20 TM-Related Promotional Points

This analysis revealed that each of the user-specific issues highlighted in the brand promotional materials was also addressed in the literature. Only three of the top 20 issues featured in the vendor promotional materials were not addressed in the literature, and these issues are either not directly related to usage (in the case of the categories “Clients of Vendor” and “Other Services Offered by the Vendor”), or are specifically focused on large organizations or language service providers (LSPs) (in the case of “Lower Translation Costs”).

Although the issues may be addressed in current literature, the focus of the brand promotional materials is quite different. The most commonly mentioned issues were technical and feature-based: file formats that the tools can process, networking capabilities, interface, and compatibility. The benefits of using TM tools (e.g. increased income, reduced translation costs, improved translation quality) appeared lower on the list. Based on the placement of the more technical issues at the top of the list, it can be assumed that vendors are not necessarily trying to convince a potential buyer of the

benefits of using a tool, but rather highlighting the value of specific features of their particular system. This may be overwhelming for a potential or new user who is not necessarily aware of the benefits of using a system and may not be informed enough to understand the intricacies of the more technical issues. This focus may also be a result of the length of time TM systems have been on the market. As these tools have been commercially available since the early 1990s, vendors may be assuming that most translators reading the materials have already made the decision to use a TM system and are now more interested in features and new developments (i.e. the more technical specifications) rather than the benefits (e.g. quality, speed).

This section has highlighted the contents of vendor advertising and promotional materials, and this analysis is further compared to the findings in other sources and in the corpus (see section 3.2.3). While it is important and necessary to review current literature, it is also important to consider what is known about the use of TMs in practice by examining current surveys and other research pertaining to TM usage.

1.6 TM-related surveys

This section describes nine surveys, conducted between 2002 and 2007, that evaluated the use of TM systems. Information about the authors, goals, methodology, dates, respondents, and findings is provided for each of the surveys in Tables 4 and 5.

Survey Distributor	Publication Date	Title
Localization Industry Standards Association (LISA) ¹³	2002	LISA 2002 Translation Memory Survey: Translation Memory and Translation Memory Standards
LISA ¹⁴	2004	LISA 2004 Translation Memory Survey: Translation Memory and Translation Memory Standards
eContent Localization Resources for Translator Training (eCoLoRe) ¹⁵	2003	eCoLoRe Translation Memory Survey 2003
Fulford and Granell-Zafra	2005 ¹⁶	Translation and Technology: A Study of UK Freelance Translators
Dillon and Fraser	2006 ¹⁷	Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption
Lagoudaki	2006	Translation Memories Survey 2006
Association of Translators and Interpreters of Ontario (ATIO)	2005	2005 Survey of Independent Translators
ATIO	2007	2007 Survey of Salaried Translators
Ordre des traducteurs, terminologues et interprètes agréés du Québec (OTTIAQ)	2006	2006 Survey on Rates and Salaries

Table 4 - Overview of TM Surveys

The surveys are addressed in chronological order in subsections, except when multiple surveys have been conducted by the same organization; in the latter cases, the surveys are grouped together. Each subsection includes an overview of the context of the survey and a discussion of the survey's findings. The survey outlines are followed by a critical comparison of the findings. Table 5 provides a summary of the administrative information regarding each survey.

¹³ The 2002 LISA Translation Memory Survey: Translation Memory and Translation Memory Standards report was written by Arle Lommel. The survey was written by Alex Lam.

¹⁴ The 2004 LISA Translation Memory Survey: Translation Memory and Translation Memory Standards report was written by Arle Lommel and edited by Rebecca Ray.

¹⁵ Both Wheatley (2003) and Höcker (2003) report on the findings of the eCoLoRe survey.

¹⁶ Although Fulford and Granell-Zafra's survey took place during late 2003 and early 2004, the findings were not published until 2005.

¹⁷ Although Dillon and Fraser's survey took place during 2004, the findings were not published until 2006.

Survey	Survey Period	Distribution Method	Distribution Group	Geographical Area	Number of Respondents
LISA 2002	Aug. – Oct. 2002	online	language professionals	worldwide	134
LISA 2004	Aug. – Oct. 2004	online	language professionals	worldwide	274
eCoLoRe	Spring 2003	online, e-mail, hard copy	ITI, ¹⁸ BDÜ, ¹⁹ industry contacts of ITI and BDÜ	United Kingdom, Germany	208
Fulford & Granell-Zafra	late 2003 – early 2004	mail	professional body of freelance translators ²⁰	United Kingdom	391 ²¹
Dillon & Fraser	July 2004 – Aug. 2004	online, e-mail	professional body of translators ²² and recent graduates and students of postgraduate translation course ²³	United Kingdom	59
Lagoudaki	July 2006 – Sept. 2006	online	promoted to discussion forums, companies and organizations, associations, educational institutions, public authorities	worldwide	874
ATIO 2005	c. 2005 ²⁴	unknown	freelance members	Ontario	193
ATIO 2007	c. 2007	online	salaried members	Ontario	119
OTTIAQ	c. 2006	online	members	Quebec	493

Table 5 - Survey Summaries

¹⁸ Institute of Translation and Interpreting, United Kingdom

¹⁹ Bundesverband der Dolmetscher und Übersetzer e.V., Germany (Federal Association of Interpreters and Translators [my translation])

²⁰ The name of the professional body to which the questionnaires were distributed is not indicated in Fulford and Granell-Zafra's article.

²¹ A total of 591 questionnaires were returned (Fulford and Granell-Zafra 2005, 8); however, as the survey was to be limited to freelance translators, the questionnaires were prescreened. This figure indicates the number of usable responses.

²² Peer Mentoring Group and French, German and Spanish e-groups of the Institute of Translation and Interpreting, United Kingdom

²³ University of Westminster cohort

²⁴ The exact survey dates for the ATIO 2005, ATIO 2007, and OTTIAQ surveys are not indicated in the reports.

1.6.1 Discussion of survey methodologies

The nine surveys listed in Table 5 encompass a range of worldwide GILT professionals; however, each survey applied a methodology to investigate specific aspects of TM usage by a target group of users. These methodologies are discussed below in terms of respondents and areas of interest.

1.6.1.1 Respondents

As shown in Table 5, respondents to the surveys ranged from freelance translators (e.g. in Fulford and Granell-Zafra, ATIO 2005), to localization professionals (e.g. LISA 2002 and LISA 2004), with most surveys focusing primarily on members of professional associations. Surveys included respondents from various locations, from international (LISA 2002 and 2004, Lagoudaki) to specific regions (Canadian provinces in ATIO 2005 and 2007 and OTTIAQ, the United Kingdom in Fulford and Granell-Zafra and Dillon and Fraser, and the United Kingdom and Germany in eCoLoRe). While the majority of the surveys were administered online or through e-mail, Fulford and Granell-Zafra's survey was mailed, and the eCoLoRe survey was also available in hard copy. Finally, the number of respondents ranged from 59 (Fulford and Granell-Zafra) to 874 (Lagoudaki), with an average of approximately 300 respondents. Despite these variations in region, sample size, and profession, a comprehensive review of these surveys provides valid insights into technology use in the field.

1.6.1.2 Areas of interest

Certain surveys were more specifically TM-focused than others. The ATIO and OTTIAQ surveys, while briefly touching on TM usage, were mainly focused on gathering other data related to their members. Fulford and Granell-Zafra's survey addressed information and communication technologies (ICT); while the survey did include TM-related questions, it is difficult to extract only the TM-related findings and draw solid conclusions on TM usage. Dillon and Fraser investigated three pre-established hypotheses, and, although the hypotheses were the main focus of the survey, the findings do address TM-related issues. The LISA surveys aimed to gather statistical data on usage (frequency, size of TMs, feature development, formats). The eCoLoRe

survey was specifically designed to address TM usage as well as training requirements and issues. Finally, Lagoudaki covered TM use in the most detail, focusing on adoption, brand usage, training, feature development, and user ratings of brands. These nine surveys collected data related to various important aspects of the field and the place of TM system usage in the GILT industries.

1.6.2 Survey summaries

1.6.2.1 LISA 2002

In 2002, the Localization Industry Standards Association (LISA) conducted its first TM survey, which set out to gauge the uptake and usage trends of TM systems in the GILT industries. Not only was this LISA's first TM survey, but it was the first large-scale survey relating to TM systems. Questions were centered on the use of TM technology, the role of open standards, and interest in exchangeability of TMs (buying, selling, or trading) amongst companies, organizations, or individuals.

Of the 134 GILT professionals who responded to the survey, the majority, 58%, were members of management (upper, middle, and lower) (Lommel 2002, 3). GILT-industry professionals (translators, writers, other employees working in the localization field) constituted 26% of the total respondents (Ibid., 5). The remaining 16% included engineers, consultants, educators, and those who did not identify their position (Ibid.).

The survey showed a high level of TM usage with almost 60% of the respondents indicating that they used a TM system for 75-100% of their work (Lommel 2002, 6). Not only was the majority using TM, but more than half indicated that they intended to extend their current usage of TM systems (Ibid., 7).

When asked in which format their TMs were stored, 90% of respondents indicated they used a proprietary or commercial format (e.g. Trados' .tmw format), 11% indicated use of an internal proprietary format (i.e. of an in-house system), and 29% used the Translation Memory eXchange (TMX) open standard "to some extent" (Lommel 2002, 11). While the majority (73%) used only one format, 27% of the respondents used a combination of formats (proprietary, internal, or TMX) (Ibid., 12). This demonstrates that one specific tool does not always meet users' needs and that often a combination of tools is used for translation activities.

As noted above, TMX was used by 29% of the respondents (Lommel 2002, 11) and, in fact, was used exclusively by 6% for storing their TMs (Ibid., 12). Despite being used by less than one-third of the respondents, knowledge of TMX was fairly widespread, with 86% of respondents indicating that they were aware of TMX (Ibid., 15). The most common reasons for TMX usage were to increase exchangeability, support open standards, and ensure greater reusability of TMs (Ibid.).

The final section of the LISA 2002 survey addressed interest in a TM marketplace (i.e. the buying, selling, and trading of TMs). At the time of the survey, there was little to no market for inter-company TM exchange (excluding client-translator TM exchange), and more than 30% of the respondents expressed no interest in the development of such a market (Lommel 2002, 16). While at least one-third of all respondents indicated interest in buying, selling, and trading TMs, respondents stated that the existence of a TM marketplace was complicated by copyright and intellectual property issues (Ibid.). Other reasons preventing further interest included quality concerns and a lack of subject overlap among companies and their TMs (Ibid.).

1.6.2.2 LISA 2004

In 2004, LISA conducted a follow-up to the 2002 survey, which aimed to expand upon and update previous findings, and was focused on TM uptake, usage, formats, and exchangeability. The questionnaire contained new questions to gather data on usage of specific commercial TM brands and areas of development of TM technologies.

The response rate for the 2004 survey was more than double that of 2002; a total of 274 responses were received (Lommel 2004, 2). As in 2002, most respondents, 62%, were members of management (upper, middle, and lower) (Ibid., 4); translators, terminologists, and other localization professionals constituted 28% of respondents (Ibid.). The remaining 10% were technical professionals (e.g. software engineers) and individuals working or studying at academic institutions (Ibid.).

The survey results showed that 59% of respondents used TM systems for 75-100% of their work (Lommel 2004, 5). In fact, 18% of these respondents used TMs for

100% of their translation work (Ibid.).²⁵ An additional 15% used TM systems for 50-75% of their work (Ibid.). Only 9% of the respondents indicated no current usage of TM systems (Ibid.). Common reasons for not using TMs included a lack of coordination among potential users within the organization, the lack of adequate support for particular file formats, a working relationship with outsourcers who did not use TM themselves, and, surprisingly, translation of hard-copy documents (Ibid., 6). According to the LISA surveys, TM usage saw neither a substantial increase nor decrease between 2002 and 2004, with 75-100% usage by 60% in 2002 and 59% in 2004 (Lommel 2002, 6; Lommel 2004, 5). Due to the structure of the 2002 survey, a comparison of the number of non-users is not possible.

Of the respondents already using TMs in 2004, 56% were planning on extending their usage, 33% were already at 100% usage, and only 11% specified no intention to expand usage (Lommel 2004, 6). As with usage in general, the number of respondents planning on extending usage remained steady between 2002 and 2004, with over half indicating their intention to increase usage (Lommel 2002, 7; Lommel 2004, 6).

The survey results show that many companies used TMs to decrease the time-to-market or improve the quality of their translations rather than to lower translation costs. (Lommel 2004, 16).

The 2004 survey also gathered information relating to TM formats. Commercial or proprietary formats were used by 84% of respondents, 8% used internal formats (i.e. those produced by an in-house system), and 30% used TMX (Lommel 2004, 10). Three-quarters (76%) of the respondents indicated use of only one format (e.g. commercial *or* internal *or* TMX) (Ibid., 11). These results are similar to those of the 2002 survey (90% commercial, 11% internal, 29% TMX (Lommel 2002, 11)) indicating relatively little change in the type of TM formats used and a negligible increase (1%) in the use of TMX.

Twenty percent of the respondents were unaware of the existence of TMX (Lommel 2004, 17). TMX usage was cited for the purposes of exchanging TMs, gaining independence from TM vendors (commercial TM systems), and facilitating the integration of additional technologies (Ibid.). Lommel notes the decrease in awareness

²⁵ The 2002 survey did not include a separate category for those who used TM systems for 100% of their work.

from 2002 and writes that “clearly more publicity about TMX and its potential uses needs to be made available to help companies understand how TMX can benefit them” (Ibid.).

The idea of a TM marketplace was further investigated in the 2004 survey. The results showed 24% interested in trading TMs, 21% in buying, and 15% in selling (Lommel 2004, 16). Although 25% of the respondents expressed no interest in buying, selling, or trading TMs, almost half of the respondents (49%) noted that they needed to investigate the issue in more detail (Ibid., 17). As in 2002, the legalities of TM exchange were cited as the strongest deterrent to the creation of a TM marketplace (Ibid.). This decrease in interest from 2002 may be indicative of respondents’ awareness of the complexities involved in TM exchange. Lommel states that “it is difficult to see how these issues could be overcome to allow any large-scale development of a TM marketplace” (Ibid.).²⁶

New to the 2004 survey was a section focused on brand data, which aimed to investigate how many and which particular brands of TM systems were being used. Over 50% of the respondents indicated use of more than one tool, with an average of three tools being used by each respondent (Lommel 2004, 11). Respondents were using multiple tools due to client requirements and the optimization of certain tools for certain tasks (Ibid., 12). When asked which specific tool(s) they were using, respondents reported use of 34 different brands (Ibid.). The ten most commonly identified are listed in Table 6 (Ibid.). It is interesting to note the predominance of Trados, as it was used by 71% of TM users. This number may have increased substantially since the purchase of Trados Inc. by SDL International in 2005 (DePalma 2005, 3).²⁷

²⁶ Nevertheless, the creation of the TM Marketplace (see section 1.4.6.4) indicates that these issues have been overcome to some extent.

²⁷ Prior to this time, SDLX and Trados were separate systems. After the acquisition, SDL produced SDL Trados 2006. SDLX is now distributed as part of the SDL Trados suite.

TM Tool	Percentage of Users
Trados	71%
SDLX	28%
Déjà Vu	24%
Alchemy CATALYST	20%
STAR Transit	19%
Wordfast	18%
In-house TM system	12%
Passolo	5%
Foreign Desk	5%
OmegaT	4%

Table 6 - Ten Most Commonly Used Tools, LISA 2004

A section pertaining to new developments in TMs was also added for the 2004 survey. In this section, respondents were asked to rank a list of features and developments based on their importance for TM technology. The following list presents the technologies, issues, and developments ranked from most important to least important as identified by respondents (Lommel 2004, 18):

- open standard for segmentation (SRX),
- development of a word count standard,
- xml:tm (allowing for the embedding of TM data within XML documents),
- standardization of attribute settings (storing information about a text),
- standardization of contextual information,
- metrics transparency (disclosure of statistical and match calculations),
- support for concept-oriented terminology,
- support for localization directives (instructions for process),
- use of TM servers,
- plug-ins to content management systems (CMS),
- open repository (a public and freely available TM),
- integration with MT, and
- TM marketplace.

The results of this question reveal that issues relating to standardization of data and display formats seemed to be important to TM users, while the exchange of TM formats (open repositories and a TM marketplace) were less so. These findings increase awareness about the areas that users identify as needing development and can help focus ongoing TM technology research.

1.6.2.3 eCoLoRe

eCoLoRe was an EU-funded project created in 2002 to “remedy a ‘severe’ skills shortage” in the localization industry (Wheatley 2003, n.p.). The 2003 survey was designed and implemented in conjunction with the mandate of the project, which aimed to “create a freely accessible repository of data for training translators in the use of translation tools” (Ibid.). eCoLoRe’s survey sought to identify current TM usage, areas of use, and training needs and concerns.

A total of 208 responses were received (Höcker 2003, n.p.). Roughly half of the respondents, 48%, were freelance translators who did not work directly with an agency, while 36% of the respondents were translators working closely with an agency (Wheatley 2003, n.p.). The results showed usage of TM by the majority of respondents (74%) with 29% using TMs daily (Ibid.). Wheatley writes that “translators predominantly engaged in technical translation are more likely to use TM products” (Ibid.)

When asked why they use TMs, 75% responded that TM usage brought about productivity gains, 70% claimed increased consistency, 55% used TMs to simplify quality control, and 41% were required to use a TM system by their client(s) (Wheatley 2003, n.p.).

When asked why they do not use TMs, the non-users indicated the following reasons: initial cost (51%), need for training (38%), return on investment (34%), difficulty of learning to use the systems (34%), low awareness of systems’ potential (34%), differential payment practices (i.e. lower rates for matches found in a TM) (30%), lack of confidence (25%), limited availability (25%), and a lack of interest (23%) (Wheatley 2003, n.p.). In contrast, 40% indicated that the difficulty of learning or the return on investment were not issues preventing TM adoption (Ibid.). Again, about one-third of non-users indicated that training, cost, and payment practices were not obstacles or factors in their decision (Ibid.). These findings provide insight into reasons for non-adoption and allow stakeholders (i.e. vendors, trainers, clients/agencies) to address the issues inhibiting TM usage.

In the key section on TM tool training, more than half (54%) of TM users indicated that they were self-taught, 12% received classroom training by an independent trainer, and 3% learned about using TM tools during a university course (Wheatley 2003,

n.p.). Almost half of the respondents thought that training was too expensive, and one-third stated that there was no training available locally (Ibid.). Most respondents (62%) thought they would benefit from classroom training, 46% expressed a desire to see training materials freely available, 36% designated online or web-based training as a useful method, and 20% indicated that one-on-one training would be beneficial (Ibid.). Overall, these results show a need for training, but variations in preferred types of training indicate that further research may be needed to establish specific strategies (i.e. by whom the training should be provided for particular groups or types of users) for improving the situation.

1.6.2.4 Fulford and Granell-Zafra

In 2005, Fulford and Granell-Zafra published the results of a survey that investigated the uptake of information and communications technologies (ICT) (i.e. tools and technologies such as word processors, web publishing software, internet search engines, corpora, TMs, e-mail, websites, and business management software) among freelance translators. The focus of the survey was to investigate the range and types of tools used, considering both the business and language aspects of translation (Fulford and Granell-Zafra 2005, 4).

Computer-aided translation (CAT) tools (including TM systems) were used by 28% of respondents, a significantly lower number than in the other surveys discussed in this section. Conversely, almost half of the respondents were not familiar with these tools (Fulford and Granell-Zafra 2005, 10).

Respondents indicated that using ICT provided benefits such as time savings, higher quality services, increased effectiveness, and improved communication (Fulford and Granell-Zafra 2005, 11). Furthermore, a large number of respondents believed that their ICT usage had augmented their income (Ibid.).²⁸ The survey results indicated that although the respondents were less convinced of the benefits of CAT tools specifically, current CAT tool users had a more positive perception of the benefits of CAT tool usage than non-users (Ibid.). Fulford and Granell-Zafra concluded that the “non-adoption of translation tools was more a function of translators’ lack of awareness of, and familiarity

²⁸ The exact figure was not given in the article.

with, these tools than an active rejection decision based on thorough knowledge of the tools and their functionality” (Ibid., 12).

When asked about ICT training, 85% reported that they were self-taught (Fulford and Granell-Zafra 2005, 9). Fulford and Granell-Zafra note that the findings of the survey highlight the need to promote and raise awareness about these tools through translators’ professional bodies and that these bodies or associations should provide a forum for translators to discuss the use of these tools and learn from each other’s expertise (Ibid., 13).

1.6.2.5 Dillon and Fraser

Dillon and Fraser conducted a TM survey in 2004, which explored three hypotheses: newly qualified translators have a more positive perception of TMs than experienced translators, translators who have adopted TMs have a more positive perception of this tool than non-users, and information technology (IT) proficiency is not the major influence on perceptions of the value of TMs (Dillon and Fraser 2006, 69).

A total of 59 usable responses were received (Dillon and Fraser 2006, 72). Of the total respondents, 85% were freelance translators, 11% were in-house translators, and 4% indicated that they worked both freelance and in-house (Ibid.).

According to the responses, 52% had regular access to a TM system and were therefore classified by the authors as TM users (Dillon and Fraser 2006, 72). While the survey did not investigate specific benefits of using TM systems, the findings revealed that 44% of non-users with three or fewer years of experience indicated that they knew many translators who used TMs (Ibid., 73). From this statistic, Dillon and Fraser conclude that “these newcomers to the profession assume that most, if not all, established translators are TM users” and that these tools are essential and necessary for success (Ibid.). Non-users indicated the time and financial investment needed to learn and use TM systems as reasons for non-adoption (Ibid., 73).

In the questionnaire, respondents were asked to indicate their main source of contact with TMs. Those with ten or more years of experience specified their own work as their main contact with TMs (Dillon and Fraser 2006, 74). Respondents with four to nine years of experience listed training sessions or workshops as their main source of

contact (Ibid.). The less experienced respondents, with zero to three years of experience, were the least likely to be in contact with TMs (Ibid.).

The findings supported two of the three hypotheses: new translators do have a more positive perception of TMs, regardless of whether or not they have used them, and TM users have a more positive perception of TMs as compared to that of non-users. (Dillon and Fraser 2006, 79). Their final hypothesis, that overall information technology (IT) proficiency is not the main influence on perceptions of the value of TMs, was disproved. The study found that an individual's perceived IT proficiency is more influential when choosing to adopt TMs than initially believed (Ibid.).

1.6.2.6 Lagoudaki

Lagoudaki's 2006 survey sought to bring TM users' opinions and attitudes to the forefront by investigating which tools are used, the environment in which the tools are used, the rate at which they are used, the perceptions and practices surrounding their use, and areas for development.

Of the 874 language professionals who responded, including translators, terminologists, project managers, and revisors, a large majority (90%) were translators (Lagoudaki 2006, 9).

TM uptake proved extremely high among respondents, with 82.5% indicating that they used TMs (Ibid., 15). Based on indicated areas of specialization, there was a strong correlation between text type (technical texts) and TM usage (Ibid., 16). Not only was TM uptake high, but 65% of TM adopters used TMs for 75-100% of their work (Ibid., 20). Almost three-quarters (71%) of TM users indicated that the decision to use TM was personal, 20% were required to use TMs by their employers, and 9% were fulfilling a client requirement (Ibid., 19). Users indicated that TM tools saved them time, improved their terminological consistency, increased the quality of translations, brought about cost savings, and allowed them to easily exchange resources (Ibid., 22). It is interesting to note that 31% of those who "voluntarily" chose to use a TM system were doing so to increase exchangeability of TMs (Ibid.). This provides an interesting contrast to the apparent lack of interest in a TM marketplace expressed by respondents to both the LISA 2002 and 2004 surveys (Lommel 2002; Lommel 2004).

Within the non-user group, 89% of respondents had heard of TM tools, but were currently not using a TM system (Lagoudaki 2006, 17). The most common reason (indicated by 28%) for non-adoption was the unsuitability of TMs for the texts being translated (Ibid.). Surprisingly, 16% of the non-users had purchased a tool, but did not know how to use it! Other reasons for non-adoption included the cost, lack of time and training, and low ROI (Ibid.). Of the non-users, 71% stated that they would be willing to try using a TM system in the future, which, in Lagoudaki's view, "raises the hope of greater dissemination of the TM technology in the future" (Ibid.).

The top ten most commonly used tools are listed in Table 7 (Lagoudaki 2006, 24). More than 50% of respondents used multiple TM tools. Anywhere from one to eight, and, on average, 2.21 different tools were used by an individual (Ibid., 23). Lagoudaki noted that the number of tools used by an individual respondent increased with the length of time he or she had been using TM systems, concluding that "users achieve a certain level of sophistication in the use of TM tools over time, which renders them more aware of the limitations and weaknesses of their current tools and drives them to investigate the TM market for new solutions" (Ibid.). The use of multiple tools may also be due to an increase in the number or variety of clients, as each client may have specific and different brand requirements.

TM Tool	Percentage of Users
Trados	51%
Wordfast	29%
SDL Trados 2006	24%
Déjà Vu	23%
SDLX	19%
STAR Transit	14%
Alchemy CATALYST	8%
OmegaT	7%
Logoport	6%
Passolo	5%

Table 7 - Ten Most Commonly Used Tools, Lagoudaki (2006)

Just over half (51%) of TM users indicated that they had received no formal training on their TM system and were self-taught (Lagoudaki 2006, 19). Eighteen percent had taken a short seminar or course, 12% received training from their employer, 7% were introduced to the tools during an academic course, 7% were trained by the

vendor, and the remaining 5% received training from another source (colleagues, local software dealers, and online user forums) (Ibid.).

The survey also addressed the issue of word-processor or text-editor integration (i.e. the integration of a TM system into a third-party word processor or text editor) (see section 1.3.2.3). The findings show that 56% of respondents preferred working within Microsoft Office Word when using a TM system, 22% preferred a proprietary text editor, and 22% reported using both comfortably (Lagoudaki 2006, 30).

The final section of the survey investigated areas for development of TM technology and aimed to address the importance of these developments to both users and non-users. The following improvements were listed, as ranked from “very important” to “not very important” (Lagoudaki 2006, 29):

- handle text in embedded illustrations,
- comply with TMX and SRX standards,
- handle PDF files,
- integrate spell checkers,
- display some context for the suggested segment match,
- construct fuzzy matches from existing resources,
- view the original source text in full,
- display a preview of the completed target text,
- handle graphics,
- support all languages,
- search as you type,
- support project-specific directives,
- sort lists in alphabetical order,
- offer OCR software compatibility,
- provide content management system (CMS) plug-ins, and
- run on multiple platforms.²⁹

The results reveal that issues relating to file formats and format standardization were of significant interest to respondents, whereas integration or compatibility (e.g. OCR, CMS) issues, as well as cross-platform support, were less important.

²⁹ Lagoudaki does not indicate whether these items were part of a pre-established list, or if respondents were free to include any development they deemed to be important. However, as the results were ranked on a specific scale (a scale of one, not very important, to four, very important), it can be assumed that respondents were given a list of items to rank.

1.6.2.7 ATIO 2005

The Association of Translators and Interpreters of Ontario (ATIO) conducted a survey of independent (freelance) translator members to gather data related to languages in which the translators work, age and experience brackets, client types, rates of pay, and tools and resources used during the translation process.

A total of 860 questionnaires were distributed, and 193 responses were received (McInnis and Takla 2005, n.p.). All of the respondents were independent or freelance translators and members of ATIO; 152 were certified translators and 42 were candidates for certification (Ibid.).³⁰ The survey results were divided into two categories: translators working in the official languages (French and English) and translators working in foreign languages.

The authors wrote that the “results were somewhat unclear regarding the use of translation memory software” by translators working in the official languages (McInnis and Takla 2005, n.p.). Forty-two percent of official language translators did not respond to the question that asked whether they used TM systems, while 31% indicated that they did not use TM systems, and 27% reported use of at least one system. Within this group of TM users, 50% used Trados and 30% used LogiTerm (Ibid.).

Results were similar among the foreign language translators. Forty percent did not respond to the question, roughly 30% reported no use of TM systems, and approximately 30% indicated TM usage (McInnis and Takla 2005, n.p.). The majority (54%) used Trados, and 33% used SDL/SDLX (Ibid.). While uptake of TM systems by freelance translators in Ontario may appear somewhat low, there seem to be clear brand preferences among those who do use TM systems.

1.6.2.8 ATIO 2007

As a complement to the 2005 survey of independent translators, ATIO conducted a survey of salaried translators in 2007. This survey had similar goals to those of the previous survey, including collecting data relating to age, experience, education, language combinations, salary, and tools used during the translation process; however,

³⁰ McInnis and Takla’s report states that 193 questionnaires were returned; however, the sum of their respondent breakdown subtotals is 194.

the survey was refocused to gather more specific data relating to salaried translators, such as type of employer and industry.

Four hundred and forty-three salaried translators were invited to complete the survey; 119 responses were received (ATIO 2007, n.p.). All of the respondents were salaried translators. Almost half (44%) of respondents indicated TM usage. The most popular tools used were MultiTrans, Trados, and LogiTerm (Ibid.).³¹

1.6.2.9 OTTIAQ

OTTIAQ conducted a survey to explore its members' professions, experience, language combinations, rates and salaries, gender, age, and use of TMs (Gauthier 2006, 1).

A record total of 493 members responded to the 2006 questionnaire (Gauthier 2006, 1). The respondents included translator, terminologist, interpreter, and student members of OTTIAQ (Ibid.). The vast majority of respondents, 93%, were certified translators (Ibid.). Over half of the respondents (60%) were freelance translators, 9% worked for a translation business, and 28% were employees in a translation department (Ibid., 2).

The 2006 survey reported a steady increase in TM usage, with uptake climbing to 45.6% from 37.2% in 2004, which was up from 25% in 2002 (Gauthier 2006, 10). Furthermore, the results show that both the mean and median annual incomes of those using TMs were several thousand dollars higher than those of non-users (Ibid.). This difference cannot solely be attributed to TM adoption, as income is without doubt affected by experience, text type, text characteristics, and many other factors; however, further investigation of these factors may provide insight into the type of translators who choose to use TM systems and the direct impact of this choice on their income.

³¹ The exact percentages for brand usage were not provided in the report.

1.6.3 Comparison of TM surveys

1.6.3.1 TM usage

Survey	Percentage of Respondents Using TMs
LISA 2002	$\geq 92\%$ (but $< 100\%$) ³²
LISA 2004	81%
eCoLoRe 2003	64%
Fulford and Granell-Zafra 2004	28% (all CAT tools) ³³
Dillon and Fraser 2004	52%
Lagoudaki 2006	82.5%
ATIO 2005	$\geq 27\%$ (but $< 69\%$) ³⁴
ATIO 2007	44%
OTTIAQ 2006	45.6%

Table 8 - Overview of TM Usage

TM uptake is reported in all nine surveys, with five of the surveys reporting more than 50% of respondents as users (Lommel 2002; Lommel 2004; Wheatley 2003; Dillon and Fraser 2006; Lagoudaki 2006). The results from the 2002 and 2004 LISA surveys show that TM usage remained steady, with 60% using TMs for 75-100% of their work in 2002 and 59% in 2004 (Lommel 2004, 5). The 2006 OTTIAQ survey highlighted a steady increase in usage from 25% in 2002 to 45.6% in 2006 (Gauthier 2006, 1). The findings from four surveys also suggest a correlation between text type (i.e. more technical or commercial texts) and TM usage (Lommel 2002; Lommel 2004; Wheatley 2003; Lagoudaki 2006).

A comparison of TM uptake and type of employment (i.e. freelance translator working directly with an agency, freelance translator not working directly with an agency, company owner, company employee) does not reveal significant differences: the findings of the ATIO 2005 survey of freelance translators indicated that approximately one-third of respondents used TM systems (both foreign and official language translators) (McInnis and Takla 2005, n.p.) compared to 44% of the salaried translators who used TM

³² The LISA 2002 survey findings do not specifically indicate the proportion of non-users; the data gathered shows that 8% of respondents use TMs for 0-25% of their work (Lommel 2004, 5).

³³ Fulford and Granell-Zafra's survey included TMs in the category of "Translation creation" software (TMs and machine translation), and further under the term "Computer-aided translation (CAT) tool" (Fulford and Granell-Zafra 2005, 10). Therefore, the results are inconclusive as to the specific number of TM users.

³⁴ McInnis and Takla state that the responses for this question are unclear. While 27% are TM users and 31% are non-users, 42% of respondents did not respond to this question (McInnis and Takla 2005, n.p.).

systems (ATIO 2007, n.p.); Lagoudaki's findings indicated that 92% of company owners were TM users, 84% of company employees used TM systems, and 81% of freelance translators were using TM systems (Lagoudaki 2006, 15). Further comparisons are not possible due to the breakdown of the findings in the remaining six surveys.

Both the LISA 2004 and Lagoudaki surveys investigated the number of different tools used by respondents. In each case, the results showed not only that translation professionals are using TM systems, but also that many of them are using more than one system. More than half of the TM users owned and used more than one tool (Lommel 2004, 11; Lagoudaki 2006, 23). The correspondence between the surveys is further reinforced, as each found that TM users employed up to eight different tools (Lommel 2004, 11; Lagoudaki 2006, 23). From a more in-depth analysis of the results and cross-referencing of the responses, it was concluded in both surveys that the longer respondents had been using TMs, the more likely they were to use multiple systems (Lommel 2004, 11; Lagoudaki 2006, 23). This may indicate that no one system met every user's specific needs, and a further investigation of the reasons behind multiple tool usage may provide hints for developers to improve their current releases. A comparison of these findings to those related to file formats allows for an interesting observation. While the use of proprietary formats can make compatibility among TM tools complicated or impossible, commercial or proprietary formats were the most commonly used (by 84% of TM users in 2004 and 89.5% in 2002 (Lommel 2004, 10)).³⁵ Only 27% and 30% of TM-adopters used TMX or open standard formats in 2002 and 2004 respectively (Lommel 2002, 12; Lommel 2004, 10) perhaps indicating that if TMs are being used in multiple systems, users are either finding ways other than open standards to achieve compatibility or dividing their resources among tools.

1.6.3.2 Tool brands

Only two surveys (Lommel 2004; Lagoudaki 2006) performed an in-depth analysis of the specific brands of TM systems being used by respondents. The results are compared in Table 9.

³⁵ The Lommel (2002) and (2004) reports indicate different percentages for format usage in 2002: 90% commercial, 11% internal, and 29% TMX (Lommel 2002, 11) versus 89.5% commercial, 12.3% internal, and 27.2% TMX (Lommel 2004, 10).

Tool Name	LISA 2004 Ranking	Lagoudaki (2006) Ranking	Difference in Position
Alchemy CATALYST	4	7	-3
Déjà Vu	3	4	-1
Foreign Desk	9	-	-
In-house System	7	-	-
Logoport	-	9	-
OmegaT	10	8	+2
Passolo	8	10	-2
SDL Trados 2006	-	3	-
SDLX	2	5	-3
STAR Transit	5	6	-1
Trados	1	1	-
Wordfast	6	2	+4

Table 9 - Comparison of Top Ten Tools, LISA 2004 and Lagoudaki (2006)

It is interesting to note the difference in placement in the lists of Wordfast (developed in 1999) and the open-source TEnT OmegaT (released in 2001); this may be a reflection of the type of TM users who responded to the surveys; the majority of respondents to the LISA 2004 survey (62%) were members of management and 28% were translators, terminologists, or localization professionals (Lommel 2004, 4), whereas 90% of respondents to Lagoudaki's survey were translators (2006, 9). Therefore, the difference in position of OmegaT and Wordfast may be a reflection of the different TM needs of the respondents (i.e. organizational use or individual use). Furthermore, as OmegaT and Wordfast are among the least expensive TEnTs and are also cross-platform, their ranking in Lagoudaki's list may indicate that they are more popular among independent or freelance translators than organizations or companies. Lagoudaki notes that the ranking of OmegaT could also "mean that the TM tool has reached a certain level of quality that has allowed it to compete with other commercial tools and earn a high standing in terms of the users' preference" (2006, 26).

The lower ranking of the localization tools Alchemy CATALYST and Passolo in Lagoudaki's list may be due to the difference in specialization of respondents to the LISA 2004 and Lagoudaki surveys. The positions of these tools in the LISA 2004 survey results may be a reflection of the higher concentration of localization professionals who responded to the survey, as LISA is a professional association within the localization industry.

The 2005 and 2007 ATIO surveys indicated Trados, LogiTerm, SDL/SDLX (McInnis and Takla 2005, n.p.), and MultiTrans (ATIO 2007, n.p.) as popular tools. The appearance of LogiTerm and MultiTrans, both developed in Canada, in the responses of the ATIO questionnaires may indicate that Canadian translators are more aware of Canadian systems and may even prefer to use these tools.

1.6.3.3 Non-adoption of TM systems

Similarities among surveys relating to the non-adoption of TM tools also become evident when results are compared. For instance, the financial and time investments are cited as deterrents in three separate surveys (Wheatley 2003, n.p.; Dillon and Fraser 2006, 73; Lagoudaki 2006, 17). Although low-cost TEnTs are available and seem to be popular among the respondents to Lagoudaki's survey (see section 1.6.3.2), price continues to be a significant factor in TM adoption. Another recurring issue preventing adoption is a general lack of awareness of the tool's benefits and functioning (Wheatley 2003, n.p.; Fulford and Granell-Zafra 2005, 12; Lagoudaki 2006, 17). Again, these findings offer insight into areas that need further exploration and attention in order to increase awareness about TM systems and provide translators with the knowledge needed to make informed decisions.

1.6.3.4 Training

Surveys which addressed training reported high proportions of "self-taught" TM users: 54% (Wheatley 2003, n.p.) and 51% (Lagoudaki 2006, 19). Fulford and Granell-Zafra's findings show not only that respondents were self-taught CAT users, but that, in fact, 85% were self-taught on all ICT (Fulford and Granell-Zafra 2005, 9). These rates demonstrate the need to increase the availability of training, in the classroom during translator training programs, through professional associations, by companies employing translators, or by the vendors themselves. Although the eCoLoRe survey (Wheatley 2003) did investigate this area in more detail, further exploration is needed to identify the training needs of potential and current TM users.

1.6.3.5 New developments in TM technology

Two surveys, LISA 2004 and Lagoudaki (2006), asked respondents to indicate areas for development of TM technologies. Despite the two year gap in survey dates, some technologies did appear in both reports: segmentation standards compliance, presentation of contextual information for matches, support for project directives, and plug-in capabilities for content management systems (CMS) (Lommel 2004, 18; Lagoudaki 2006, 29). The appearance of these items in the results of both surveys may indicate that little progress had been made between 2004 and 2006, especially in regards to SRX support, which was indicated as the most important development in the LISA 2004 survey and the second-most-important development in Lagoudaki's findings (Ibid.). The desire to see contextual information for matches (indicated as fifth-most-important development in both surveys (Ibid.)) may indicate that current users are primarily using tools that have adopted the SB approach to TM storage and retrieval, rather than the CSB approach (see section 1.3.2.5.1). While some TM systems using the CSB approach were on the market at the time of both surveys (e.g. MultiTrans), awareness among TM users of their existence and functioning may have been low.

While surveys of translators and TM users clearly provide valuable data, as seen in this review, they are not the only means of exploring opinions about TM tools (see section 2.1 for a discussion of possible methods for obtaining translators' viewpoints). The use of data from mailing lists is another option; two investigations of this nature are discussed below.

1.7 García and Lantra-L

García took a different approach to examine the attitudes and opinions of TM users (García 2003; García 2006), centering his investigations on Lantra-L, a mailing (distribution) list.³⁶ As of July 1, 2003, Lantra-L had 1148 subscribers from 66 countries and distributed an average of 800 messages per week (García 2003, n.p.).

³⁶ A mailing or distribution list is a collection of e-mail addresses. Usually mailing lists are subject- or topic-specific. Users can subscribe to the list, automatically receive new messages from other members, and send messages of their own.

García notes that “by examining what TMs Lantra-L professionals are talking about (and indeed, not talking about) we discover that apart from objective views on different brands – principally Trados, DéjàVu and Wordfast – translators who have already made their choice rally around their standard and defend it with vigor” (García 2003, n.p.). For this study, García examined 772 “content-rich messages” made by 134 different contributors (Ibid.). He then identified, if possible, the preferred TM brand of each contributor and performed targeted searches for these brand names in the Lantra-L archives (Ibid.).³⁷ From there, García was able to synthesize and analyze the issues discussed relating to each brand. Topics of discussion included filters for file format conversion (section 1.4.2.1), terminology management (section 1.2.2), sub-segment matching (section 1.4.3.1), technical support (section 1.4.4.2), and price (section 1.3.3.2).³⁸ Furthermore, he highlighted the extent to which users cross over from tool to tool, indicating when and how the tools are used in a complementary fashion or as substitutes for one another (e.g. as seen in the headings: “Trados vs DéjàVu”, “DéjàVu as Trados substitute”, “Trados plus DéjàVu”, and “Wordfast versus Both” (García 2003)).

In a later examination, García comments on responses received to a particular question posed to Lantra-L members: “Could you explain in a few words what you actually gain by using them [TM systems]?” (2006, 97). By comparing this information to the data collected in his previous study (2003), García was able to make four observations about translation tools: two benefits and two drawbacks to TM use. Firstly, learning a TM system involves a steep learning curve (see section 1.4.1.1). Secondly, productivity is increased when using a TM system, especially for those who translate repetitive texts (see section 1.4.5.1). Thirdly, the use of TMs for non-repetitive texts may also be a worthwhile investment (see section 1.3.1.1). Finally, payment issues (i.e. receiving less money for fuzzy or exact matches) can undercut the translators’ financial and productivity gains (see section 1.4.6.2) (García 2006, 98).

By examining the viewpoints and opinions on Lantra-L, García was able to identify and comment on the issues surrounding TM usage; this approach and its results

³⁷ The Lantra-L message archive is available online to browse or search at: <http://segate.sunet.se/cgi-bin/wa?A0=LANTRA-L>.

³⁸ Note that García’s investigation, while focused on TM systems, also included issues relating to other TEnT components.

differ from those of the surveys discussed in the previous section. While the surveys revealed uptake and usage trends based on statistics, García's examination revealed the reasoning and motivations behind the decisions made by users.

Subscription to a mailing list allows TM users to express themselves freely outside of the constraints often imposed by surveys. Although membership to the mailing list Lantra-L is significant, the volume of information received by a member (an average of 800 messages per week! (García 2003, n.p.)) may be a deterrent to joining. Furthermore, the messages received may range in subject from Russian translations to Spanish terminology, or political jokes and comics to theoretical discussions. The Lantra-L member is committed to receiving these e-mails and discussions relating to all aspects of translation and must sift through the contents to identify relevant messages.³⁹ While both of García's studies offer an interesting examination of the attitudes and perceptions of translators, the 2003 study solely investigated what was being written about three specific brands, and the 2006 study was based on responses to a specific question. For these reasons, further investigation of TM users' thoughts and discussions is warranted.

1.8 Concluding remarks

This chapter has described TEnT components, considerations in choosing to use a TM system and deciding on a particular brand, and the issues surrounding the use of TM systems. The summary and analysis of current TM-related literature, surveys, vendor promotional materials, and mailing list examinations have contextualized TMs within the industry and explored certain aspects of usage. However, it is evident that a deeper understanding of the issues surrounding TM usage – particularly as seen from the viewpoint of language professionals themselves – is needed. The variety of issues presented in this chapter (from technical considerations, such as file formats, language support, and operating systems to situational considerations, such as client relations and training) and the organization of these issues into a hierarchical structure has provided the necessary framework for further investigation.

³⁹ It is possible to receive the messages in “digest” format, that is one single e-mail containing all messages sent within the previous 24 hours (Niska 2009).

I believe that a broader examination can be carried out through the analysis of postings in online discussion forums, which are online areas where people with a common interest can seek support, help, and advice, encourage one another, and commiserate. An analysis of discussion forums would allow me to collect topic-specific data, and observe translators' unprompted thoughts and opinions. By analyzing forums relating specifically to TM systems – but not necessarily to particular brands, which might bias the results – I believe I will be able to analyze a wide variety of spontaneous contributions from translators. I will discuss my methodology for choosing this primary source and creating a corpus of translators' opinions in Chapter 2.

Chapter 2 – Methodology

In this chapter, I will discuss the steps taken to indentify a primary resource, design and create my corpus, and classify the corpus data.

2.1 Choice of primary source

In order to gain access to translators' attitudes towards and opinions of TM systems, a primary source was needed. Numerous forms of primary sources were possible, including TM user interviews (section 2.1.1), a topic-related survey (section 2.1.2), the analysis of mailing lists such as Lantra-L (section 2.1.3), or an investigation of online discussion boards (section 2.1.4). In this section, each of these options is outlined and discussed.

2.1.1 TM user interviews

Interviews provide the most direct method of obtaining information from a specific group of people relating to a specific topic. Although one-on-one interviews could provide the type of data necessary for an investigation of TM perceptions and attitudes, this method would pose numerous difficulties within the scope of the current project. The aim of this research is to examine numerous users' opinions, and the amount of time needed to conduct a sufficient number of interviews did not fall within the restricted timeframe for this project. Furthermore, a geographical restriction would have to be imposed due to travel and/or long-distance phone call expenses. Finally, interviews generally follow a question-and-answer format, thereby reducing the opportunity to obtain unprompted feedback. For these reasons, the TM user interview method was not a practical or feasible choice for a primary source.

2.1.2 TM usage survey

Surveys also allow for the collection of specific information from a limited sample group. Numerous surveys relating to TM use have been conducted, although the majority have been focused on statistical data (see section 1.6). Surveys offer many advantages: they can be inexpensive to administer, they may cover a large sample or a focused respondent base, they may be administered remotely or locally through a variety

of methods (e.g. mail, online, telephone), and they can increase or ensure standardization of data. However, surveys can be relatively inflexible: they are based on specific questions which elicit somewhat restricted or highly focused responses from a pre-defined respondent base and may not leave room for unexpected or unanticipated responses. Furthermore, response rates are not predictable and vary from project to project.⁴⁰ Finally, the factors motivating survey completion are external (e.g. obligation, interest in the topic), rather than internal. The respondent is being prompted to state his or her opinion in relation to one or more specific questions or topics rather than choosing to express a spontaneous or unsolicited view. While a survey could provide valuable insight into translators' perceptions of TM systems, I am more interested in complementing the existing survey data by collecting and analyzing users' unprompted opinions. For these reasons, I decided to exclude a survey as a primary source.

2.1.3 Mailing lists

As discussed in section 1.7, the analysis of mailing list messages is an unobtrusive method for accessing translators' opinions. Mailing lists are topic-specific, and one such as the translation-related Lantra-L list, could provide the type of data necessary for this research.⁴¹ However, in order to access the data contained in the mailing list, one must subscribe or perform focused searches through the archives. Lantra-L is not specific to translation technologies, and messages cover a wide range of topics relating to translation theory and practice. A subscription would not guarantee that a sufficient amount of data could be obtained within the time constraints. While the archives are accessible, it would be necessary to identify keywords or search terms in order to retrieve the TM-related content. This use of keyword searches could favour or restrict the retrieval of messages discussing TM system usage (i.e. based on keyword choice, increasing the possibility of retrieving messages discussing specific aspects of TM usage, while preventing others relating to different aspects of TM usage from being retrieved). Furthermore, as two investigations using this method and list have already been conducted, it would be

⁴⁰ The response rates for the surveys discussed in section 1.6 varied from 21% (Dillon and Fraser 2006, 72) to 35% (Fulford and Granell-Zafra 2005, 8)

⁴¹ Other translation technology mailing lists (such as those on Google Groups or Yahoo! Groups) were also considered; however, the use of these would pose the same difficulties as those associated with Lantra-L. (See section 2.2.2 for an evaluation of Yahoo! Groups as a type of discussion board.)

beneficial to use a different primary source in order to expand the current body of knowledge. For these reasons, I decided against an analysis of mailing list messages.

2.1.4 Discussion boards

The final type of primary source considered was discussion boards. A discussion board is an online area where members can ask questions and seek advice from their peers, and comment on the issues that are important to them. Discussion boards are often divided into subject-specific forums, and within a forum, the discussions are organized into threads. A *thread* begins when a user posts a comment or question on a new, specific sub-topic within the more general subject of the forum. All subsequent messages related to this original sub-topic are added to the thread, creating a chronological listing of users' comments, questions, or answers. Each message is called a *post* or a *posting* and contains the information written by a single user at one particular point in time (see section 2.1.4 for a more detailed description of discussion boards).

The analysis of postings on discussion boards would allow for the collection of topic-specific information written by translators. The information would be readily available, without having to rely primarily on keyword searches. Furthermore, numerous active translation-specific discussion boards are freely available on the web, providing a sufficient amount of data for collection and analysis. The collection of data from this type of source would allow me to observe translators' spontaneous comments and questions about TM systems in an unobtrusive manner.

However, the use of discussion boards as a primary source would also present numerous challenges. The type and amount of TM-related data on discussion boards is unpredictable and continually changing. In order to collect a reasonable sample, the development of a systematic method for gathering the data would be necessary. Moreover, I would need to ensure that the data collected contained information on a wide variety of topics related to TM-usage written by a range of contributors. The method for data collection would have to be designed to maximize this variety without preventing the collection of a random sample. Finally, the use of discussion boards as a primary source would involve the analysis and annotation of free text (inevitably more challenging than interpreting responses to closed-ended questions) and would require the

development of a classification system to ensure consistency. The development of a systematic method and classification system as well as the collection and analysis of the data were within the scope of the current project.

The validity of discussion boards as a primary source is reinforced by the indication of discussion forums as a frequently consulted resource in three surveys examined in section 1.6: eCoLoRe (Wheatley 2003, n.p.; Höcker 2003, n.p.), ATIO 2005 (McInnis and Takla 2005, n.p.), and Fulford and Granell-Zafra (2005, 10). In the eCoLoRe survey, 21% of respondents indicated they wished to have contact with other TM users and were planning to investigate or had already investigated user groups for support (Wheatley 2003, n.p.). The results from Fulford and Granell-Zafra's survey indicated that roughly one-third of the respondents participated in electronic mailing lists or online discussion groups (2005, 10). Moreover, 33% used online translation marketplaces such as ProZ.com, TranslatorsCafé.com, and aquarius.net for marketing and work procurement, and just over half of the respondents were familiar with these sites (Ibid.). The ATIO 2005 results revealed that translators' forums were widely used resources by both official and foreign language translators. McInnis and Takla write, "This indicates that independent translators value the opinion of their peers and rely on their expertise" (2005, n.p.).

For the reasons outlined above, and despite the expected challenges, I decided that discussion boards would be the best choice for a primary source, allowing me to readily access translators' opinions while avoiding some of the drawbacks inherent in the other methods. Before examining specific discussion boards, the next step was to verify that I would be ethically able to use the contents of discussion boards in my analysis.

2.1.4.1 Privacy and discussion boards

Before finalizing my decision to use discussion boards, I needed to ensure I could ethically use discussion board content without violating the privacy of any of the authors. I discovered that the use of information posted on discussion boards is not considered to violate the author's privacy. In an article entitled "Electronic Eavesdropping: The Ethical Issues Involved in Conducting a Virtual Ethnography", Clegg Smith cites the research conducted by the Project H Research Group, an international team of scholars who

discussed the ethics involved in online research. Project H did not obtain permission for the analysis of publicly posted messages. Clegg Smith writes:

In making this decision, Project H set an important precedent, and they justified their decision as follows, ‘We view public discourse on Computer Mediated Communication as just that: public. Analysis of such content...is more akin to the study of tombstone epitaphs, graffiti, or letters to the editor. Personal? Yes. Private? No.’ (Clegg Smith 2004, 230)⁴²

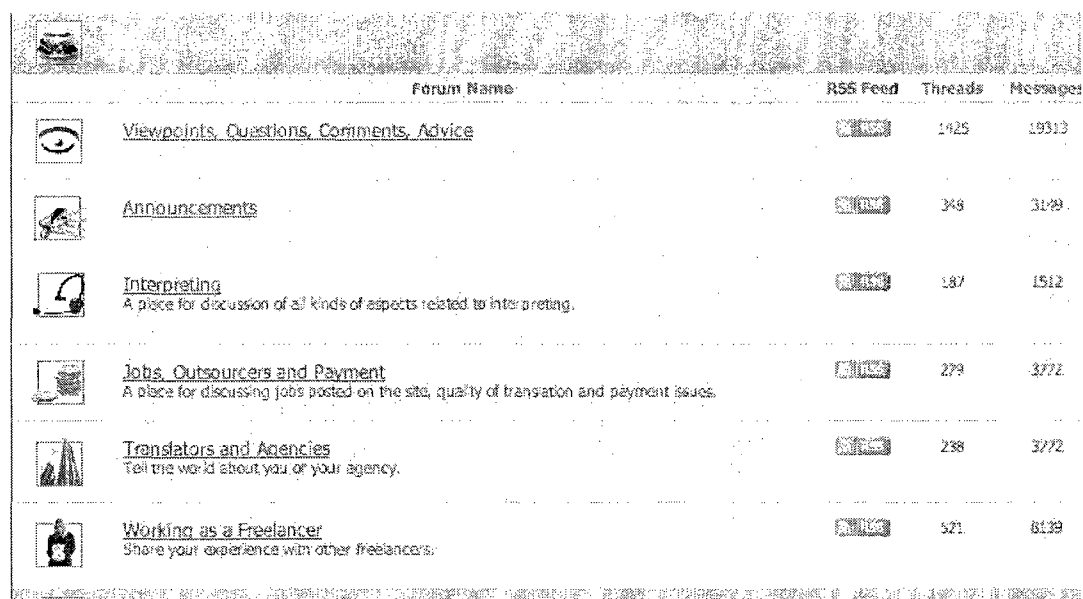
The privacy policies of two well known translation-related forums, ProZ.com and TranslatorsCafé.com (see sections 2.2.5 and 2.2.6), were also examined. The ProZ.com site recommends “that anyone using the Internet assume that anything they post online will be viewable by all people for the rest of time” (ProZ.com 2009b). Furthermore, the ProZ.com privacy policy states that a member’s username, and not necessarily his or her full name, will appear in public areas of the site, thereby ensuring that members’ privacy is protected (ProZ.com 2009c). The TranslatorsCafé.com discussion board rules of conduct state that “persons using materials from this site shall obtain permission from the author or *quote the materials with credit to the author and with a link to TranslatorsCafé.com*” (TranslatorsCafé.com 2009b) [emphasis added]. An investigation of these sites’ privacy policies and of previous research of this type thus revealed that the information posted on discussion boards is available for use when properly referenced.

2.1.4.2 Terminological clarifications

Before outlining the discussion board selection process, some terminological clarifications are necessary. As noted previously, discussion boards are often sub-divided into smaller, topic-specific forums, and these forums are often further categorized into specific groups. It should be noted that the term *forum* can be used in two ways. Firstly, *forum* can generally refer to an online area where users, usually registered members, discuss a topic or many topics. Secondly, *forum* can refer to a designated, topic-specific collection of discussions. For the purposes of this research, the term *discussion boards* is used for the more general meaning, an online area for discussion, and the term *forum* is

⁴² Note that Clegg Smith is not alone in this regard. These views are in line with approaches taken by other researchers conducting studies of data collected via computer-mediated communication over the past decade, including Herring (1996), Eysenbach and Till (2001), and Bowker and Herrera (2004).

used to refer to specific collections of discussions relating to a particular topic. Figure 1 shows the specific forums within the “General Discussion” category on the TranslatorsCafé.com discussion board.



	Forum Name	RSS Feed	Threads	Messages
	Viewpoints, Questions, Comments, Advice		1425	19313
	Announcements		348	3129
	Interpreting A place for discussion of all kinds of aspects related to interpreting.		187	1512
	Jobs, Outsourcers and Payment A place for discussing jobs posted on the site, quality of translation and payment issues.		279	3772
	Translators and Agencies Tell the world about you or your agency.		238	3772
	Working as a Freelancer Share your experience with other freelancers.		521	8139

Figure 1 - Discussion Board Organization

The threads within a forum are usually displayed in chronological order, with most recently active (i.e. most recently posted-on) threads appearing at the top. A thread may be labeled as *sticky*, bumping it to the top of the list where it will remain until this label is removed. Sticky threads are often the most active or posted-on threads and contain important or often-sought information. Figure 2 shows a listing of threads in the “CAT Tools Technical Help” forum on ProZ.com.

	Topic	Poster	Replies	Views
	Tools for swapping language combination in a created TM?	TRIDIOM Jan 22, 2008	7	562
	CAT Tool to translate Guarani	Kyju Jan 18, 2008	4	766
	Viewing segments in a Trados TM	Ⓜ Noha Kamal Jan 13, 2008	1	620
	To write with Cyrillic	Francesco Damiani Jan 9, 2008	4	1150
	A freeware tool for translation of LaTeX and other plain text files	Aleksandr Okunev Jan 3, 2005	10	3331
	Which CAT tool for Microsoft Publisher files (.PUB)?	J-a-n S-ndstr-m Dec 18, 2007	4	1205
	What CAT tool(s) to use?	Ⓜ kyanzes Dec 12, 2007	8	1531
	Which CAT-tool can do this type of propagation besides ForeignDesk? (□ 1... 2)	Ⓜ Katalin Horvath McClure Dec 13, 2007	16	2283
	How to Lock/protect variables in XML	Ⓜ Andovar Pte Ltd Oct 10, 2007	3	1082
	What CAT tool to use for Corel Draw v11/12?	Ⓜ Marcus Malabad Oct 13, 2005	9	3909
	Advice on CAT Tools	Jacy Dec 4, 2007	4	1053

Figure 2 - Threads within a Forum

A forum may have anywhere from zero to an unlimited number of threads, and a thread may contain anywhere from one to an unlimited number of postings. The timeline, or lifespan, of a thread may be very short (e.g. one day) or very long (e.g. five years). It is important to note that a thread is never truly inactive; users may continue to post on the thread as long as it remains open (i.e. not locked (see below)) and visible in the forum. In this sense, discussion forums are extremely dynamic, and their content may change from one minute to the next. The postings in a thread create a discussion, a series of replies and responses to other message authors' ideas or opinions. Figure 3 shows three postings on a thread entitled "What is a TM?" posted in the "CAT Tools Technical Help" forum on ProZ.com.

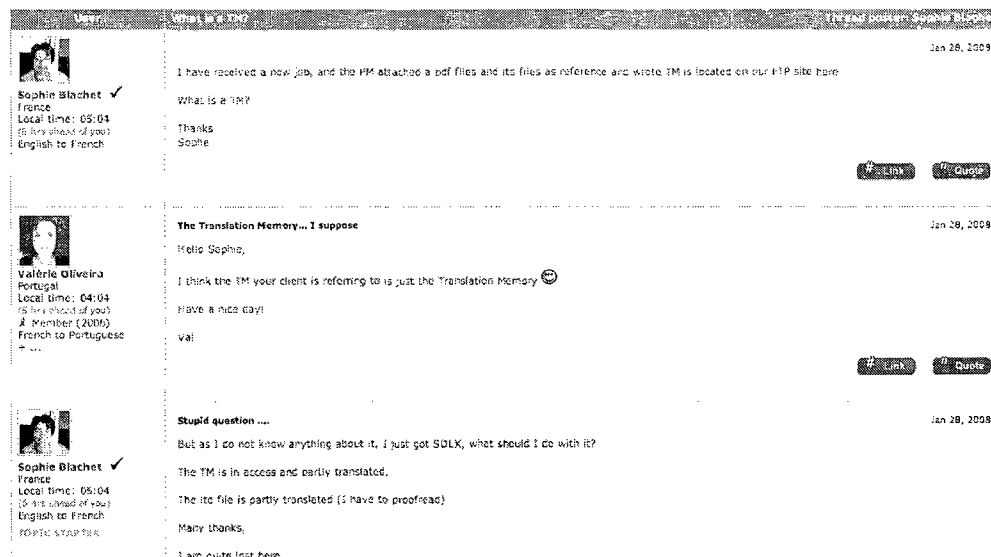


Figure 3 - Postings

The organization of large discussion boards may be best described through the use of an example. At TranslatorsCafé.com, the discussion board is divided into six categories. The categories are labeled “General Discussion”, “Translation”, “Other Things We Can Do...”, “Special”, “Language-Specific Forums”, and “TranslatorsCafé.com” and contain a total of 54 forums (TranslatorsCafé.com 2009c). As of May 8, 2009, the largest forum was “CAT and other Software for Translators” (located in the “Translation” category), which was comprised of 683 separate threads containing a total of 4337 posts (Ibid.).

Forums are usually monitored by a staff member or volunteer called the moderator. A moderator is responsible for ensuring that the content of the thread is related to the specific topic of discussion and that a post does not violate the posting rules of the discussion board. Examples of violations may include posting advertisements or spam, posting the same message more than once, using foul or offensive language, or discussing illegal activities (TranslatorsCafé.com 2009b; ProZ.com 2009a). A moderator may *freeze* or *lock* a thread preventing further posts if the content is in violation of the site rules. In addition to freezing threads, the moderator also decides if a thread should be marked as sticky and when sticky threads should be “unstuck”. Finally, the moderator is responsible for relocating off-topic threads to the correct and appropriate forums, thereby ensuring that the forums contain only threads which directly relate to the topic of the forum.

2.2 Discussion board selection

This section provides an overview of the investigation of six different discussion boards to assist in choosing specific sources for the corpus. In order for a discussion board to be considered for use as a source, it needed to meet certain requirements as outlined in the *Scope and Limitations* in the introduction (section 0.5). The board must have *substantial and active membership* to ensure that a wide variety of attitudes towards tools are available for collection. The board must have *international representation* (i.e. no specific geographical focus) to ensure the attitudes expressed are not solely those of translators working in one specific region, as the issues related to TM usage may be different in certain countries (e.g. price of the tool, technical support, language support). Furthermore, the board must *not be targeted towards translators working with specific language pairs* (e.g. English into French), as authors might be then discussing issues only related to those specific languages, and the purpose of this analysis is to gather opinions of translators in general. As English is my first language and is often the *lingua franca* used by a group of people with diverse linguistic backgrounds, the board must have a *high volume of English postings* to provide sufficient English data for inclusion in the corpus. Moreover, the board must contain an adequate number of *technology-related forums not exclusively or primarily organized by brand name* to eliminate the possibility of only obtaining brand-specific data, thereby ensuring the opinions reflect issues related to a variety of brands. Finally, it is important to retain *more than one discussion board* as a primary source in order to ensure the postings were made by a wide variety of authors from numerous different geographical locations and online communities.

2.2.1 English Spanish Translator Org

English Spanish Translator Org is a discussion board encompassing all aspects of translation, specifically English-Spanish and Spanish-English translation. It contains almost 60 individual discussion forums ranging from “Sports” and “CAT Tools Troubleshooting” to “Spanish to English Medical Translation” (English Spanish Translator Org 2006a). As of May 8, 2009, the site had 250 registered members (English Spanish Translator Org 2006b). Although the site had a significant membership base and posting volume, it was excluded as an option because it was focused on one specific

language pair. Furthermore, a number of the postings were in Spanish, and, as the corpus was to contain only English postings, these postings would be excluded, thereby reducing the number of usable postings.

2.2.2 Yahoo!

Yahoo! Groups is an online community “where people with a shared interest meet, get to know each other, and stay informed” (Yahoo! 2009b). Members can join groups and post their questions, comments, and opinions. Group members can view these postings in one of four ways: individual e-mails (one e-mail per post or message), a daily digest (messages delivered in one e-mail daily or in batches of 25), special announcements (only moderator-posted special announcements), or web only (no e-mails, only web access) (Yahoo! 2009a). While “web only” membership satisfies the requirements for inclusion as a discussion board, a quick search of the term “translation memory” reveals that the majority of TM-related groups are brand-specific (Yahoo! 2009c). Of the first ten results, eight were brand-specific (e.g. MetaTaxis, Swordfish, Transit, Wordfast) (Ibid.). Consequently, in order to retrieve data that did not only relate to specific brands, Yahoo! Groups could not be considered as an option for analysis.

2.2.3 Global Languages

Global Languages is a European translation vortal and offers a discussion board for registered users.⁴³ A preliminary investigation revealed that the four forums contained only four postings total, three of which were in French, and none of which related to TM systems (GlobalNet Services 2009). For these reasons, the discussion board met neither the language nor volume requirements for inclusion and was eliminated as an option for analysis.

⁴³ A vortal, or vertical portal, is an industry-specific web portal and often offers relevant articles, news topics, “community and collaboration capabilities, and e-commerce services for products and services” in the industry (National Library of Australia 2007).

2.2.4 aquarius.net

The translation vortal aquarius.net was launched in February 1995 (aquarius.net 2009a). Its discussion board, or “answers” section, lists six forums including a “computer-related questions” forum (aquarius.net 2009b). However, when a query of all questions in this forum was made, only one thread was visible (Ibid.).⁴⁴ Therefore, the discussion board at aquarius.net did not meet the volume requirements and was eliminated as an option.

2.2.5 TranslatorsCafé.com

The translation vortal TranslatorsCafé.com was launched in 2002 with the purpose of providing a directory of translation agencies (TranslatorsCafé.com 2009a). The discussion board was established in early 2003 (Ibid.), and as of May 8, 2009, had 90,996 registered members (Ibid.). While the site does offer language-specific forums, all other forums are exclusively in English (TranslatorsCafé.com 2009c). Membership is required to post; however, anyone may view the contents of the discussion forums. On May 8, 2009, the non-brand-specific discussion forum “CAT and other Software for Translators” had a total of 4337 messages, providing sufficient volume for analysis (Ibid.). For the reasons outlined above, TranslatorsCafé.com met the requirements and was retained as a data source.

2.2.6 ProZ.com

The translation vortal ProZ.com has an extremely active discussion board. The board contains over 125 forums, some of which contain more than 1,000 threads (ProZ.com 2009d). The categories include language-specific forums, tool-specific forums based on brand name, and general forums related to working as a freelancer, running a business, and using CAT tools (Ibid.). Although no specific membership statistics were available, the sheer volume of posts indicated significant and regular activity. With the exception of the language-specific forums, all posts are in English. Finally, the existence of forums such as “CAT Tools Technical Help” and “Smart

⁴⁴ Note that aquarius.net recently underwent extensive reorganization. Before this reorganization, numerous computer-related postings were online; however, the archived data is not accessible.

Shoppers” implied that the use of CAT tools and all aspects related to their use were discussed in the forums, outside of the brand-specific forums (Ibid.). Therefore, ProZ.com satisfied the necessary requirements and was retained as a source for gathering data.

After investigating the above six discussion boards, I decided that TranslatorsCafé.com and ProZ.com would be the sources used for gathering the corpus data.

2.3 Sample corpus

In order to determine the amount and type of data available, a sample corpus was created. When building a corpus of discussion board postings, it is important to note that a thread is a logical progression of thought and discussion; consequently, postings cannot be analyzed in isolation, but must be examined within the context of the whole thread. Therefore, decisions on corpus inclusion and exclusion were made at the thread level. This thread-based inclusion/exclusion policy was used when creating both a sample corpus and the main corpus.

This is a relatively new approach, and it was not immediately clear how much information could be extracted from postings and how long it would take to gather and analyze an acceptable amount of information. In order to address these issues, a sample corpus of approximately 100 postings was created to aid in a preliminary evaluation of the number of postings expected to be necessary to obtain the amount of data required, and the time needed to build and analyze the corpus. In order to retrieve the postings most relevant to TMs, the data for the sample corpus was collected from the “CAT and other Software for Translators” forum on TranslatorsCafé.com. The earliest thread from this forum was created on January 22, 2003, and the inclusion of all threads from this date forward would provide an excessive number of posts for the sample corpus. In order to limit this to a more reasonable number, a date restriction was imposed: a thread needed to have been posted on between January 1, 2005 and the present (March 2009) in order to be included in the sample corpus. All other threads (threads that had not been posted on since December 31, 2004) were automatically excluded. This four-year time period was expected to provide a relatively manageable number of postings on pertinent topics

within a reasonable date range. As the forum relates to any CAT tools and software for translators, all threads exclusively related to CAT tools other than TM systems (e.g. word processors, file converters, terminology management systems, web browsers) were excluded from the corpus.

I began compiling the corpus from the earliest threads (posting closest to January 1, 2005) in the eventuality that all threads from the forum would need to be included in the sample corpus. I chose to use a Microsoft Office Excel workbook as the format for my corpus. An Excel workbook allowed me to create multiple spreadsheets within one file, a necessary feature as the project required tracking both threads and postings. Although there were no complex mathematical calculations involved, a spreadsheet allowed for the entry of data in a table format that was easy to design and display as required. Columns contained various pieces of information about the posting, and each posting was contained in one single row. Furthermore, the sorting functions allowed for quick and easy organization and reorganization of the data. A final advantage was that when populating a cell, the program proposed previously entered data based on exact character matching, thereby allowing for increased consistency of the data and reducing the number of keystrokes needed to add a posting to the corpus.

A preliminary classification system was developed based on topics and brand names mentioned in the posts. This initial classification system is not discussed in detail; the classification and layout of the main corpus are discussed later in this chapter (section 2.7.3). Figure 4 shows the contents of the sample corpus. The columns on the left contain administrative data, the large middle column contains the body of the posting, and the columns on the right represent classification categories. (Additional representations of the corpus structure are located in Appendices F and G.)

Posting title	P Post. #	TC Post. #	Date	Author	Country	Language	Content	Ag	Re	Am	A+	B+	CC	CO	CR		
							I've explained in a last conference my confusion presentation (see link below) the loss of Inbound/Inward vs Outbound/Outward. Translators don't have a clear idea of what is Inbound/Inward vs Outbound/Outward. This presentation specifically looks at various translators in several commercially available TM tools. Particular on CAT applications based on whether they are Inbound or Outbound translation focused. With a move in the direction of Inbound-type enterprise-wide translation with Translation Memory (TM) tools, it is also essential to use such an analysis for TM tools. Most TM translators I know use CAT tools (aka TM tools) for Outbound/Outward translation, yet I do know of some who are involved in Inbound-type translation jobs. Inbound/Outbound Translation. Presented in the same context. Localization is customer support, at the local level, which conference, Bonn, Germany, June 29 - July 1, 2004, by Jeff Allen. http://www.docuport.com/infocenter/program.htm Permission has been granted by the conference organizer (Multilingual Computing & Tech (ed)) to make an electronic copy of this presentation available for free. It is available at: http://www.docuport.com/infocenter/program.htm Jeff http://www.docuport.com/infocenter/program.htm										
Choosing tools based on Inbound/Outbound	TC06-01	50061	01/02/05	Jeff Allen	France	En											
RE: Choosing tools based on Inbound/Outbound	TC06-02	50111	01/02/05	Maxi Schweizer-Bauder	Canada	En, GE	Welcome to the list, and happy new year. Give a year hello to this conference, and experienced translator while Inbound, Inward, Outbound, Outward, ... refer to? I feel as though I had just stepped into a foreign country. Maxi										
							Yes, no problem explaining this concept. It is actually described in more detail in the PowerPoint slides of the presentation (converted to PDF format at the link). A very brief history of when Inbound translation emerged is explained in one of the slides. Outbound/Outward Translation (also called Translation for Publication or Dissemination) is taking a source document and translating it into one or more other languages for the purposes of publishing the target language version. This is primarily what professional translators do. Inbound/Inward Translation (also called Content or Information Gathering as well as Assimilation Translation) is translating for the purpose of understanding a document in a language that you do not know. This type of translation activity used to only comprise part of the translation market prior to 1990. It was used by the Translation Bureau at the European Commission for the rapid translation of meeting minutes and other types of verbatim, short-term information. Starting in Dec. 1997 with the first online Machine Translation system (Mistral), the Inbound/Inward translation market took off. The online translation of the Star report/Commission in 1998 is an example of this. For some translators involved with Inbound and Outbound translation activities. They might do background research using an Inbound approach with the language pair to fully complete an Outbound translation-focused job. If there is anything unclear in the PDF document at the provided link, I'm happy to explain it more.										
RE: Choosing tools based on Inbound/Outbound	TC06-03	50413	01/02/05	Jeff Allen	France	En											

Figure 4 - Sample Corpus

Ninety-nine postings were included in the sample corpus; however, not all of these postings contained TM-related content, or classifiable content (i.e. discussions of or references to the use of TM systems). Some posts contained comments such as, “thanks a lot” or “I have the same problem”; these ‘empty’ postings were added to a separate file and are referred to hereafter as *discarded postings*.

The initial classification system had 47 topic categories (e.g. payment, compatibility, networking capabilities, price), with an average of five postings per category, and 20 brand-specific categories with an average of seven postings per brand. After examining this data, I decided that approximately three to four times more data than contained in the sample was needed from each of the two discussion boards (ProZ.com and TranslatorsCafé.com) in order to obtain sufficient data for analysis and discussion, which would contain a range of different perspectives on a wide variety of tools and issues from several different authors. Based on calculations of the sample corpus data, I predicted that 23% of postings would not contain TM-relevant data and would be discarded postings. Taking this percentage into consideration, it was established that approximately 389 to 519 postings from each discussion board needed to be included in order to obtain 300 to 400 classifiable postings.

A further investigation of the “CAT and other Software for Translators” forum indicated that the forum was not topic-biased or brand-biased (it contained discussions relating to many different aspects of TM usage and many different TM systems) and contained sufficient data to satisfy the volume requirement (389 to 519 postings). However, due to the structure of ProZ.com, there was no comparable forum relating to all aspects of TM usage on the discussion boards. (The hierarchical structure of the two discussion boards is listed in Appendix C). While the “CAT Tools Technical Help” forum contained many relevant threads, these did not include discussion of issues such as the purchasing of TM systems or payment for translations done with TM systems. The relevant threads on ProZ.com were divided among many forums (e.g. “Money Matters”, “Smart Shoppers”, “Business Issues”, “CAT Tools Technical Help”). While I considered using only TranslatorsCafé.com and not ProZ.com, I decided to retain both discussion boards, as variety in the corpus was important. Consequently, the development of a systematic method for selecting forums containing TM-related threads was needed. The development and implementation of this method, completed between March and May 2009, are discussed in detail in the following section.

2.4 Forum selection

Due to the dispersion of TM-related threads among numerous forums on ProZ.com and the absence of a forum parallel to TranslatorsCafé.com’s “CAT and other Software for Translators”, a systematic method for identifying potential threads was needed. Two options were possible: the identification of keywords in order to perform a keyword search within the discussion board, which would reveal TM-related threads, or the investigation of all forums and subsequent extraction of relevant threads only. While the first method appeared to be more direct, I did not wish to use keywords for thread identification as the keywords may alter the neutrality or objectivity of the results by favouring or excluding certain topics.

The dismissal of a keyword search to identify possible threads left me with one option: investigating all forums and extracting the TM-relevant threads. A detailed, manual examination of all forums and threads was not possible due to the volume of

threads, the number of forums, and time restrictions; therefore, a systematic method was needed in order to identify the forums containing pertinent threads.

2.4.1 Board-wide keyword search

For this process, a keyword search was the clearest choice. Note that this keyword search was not being used to identify the *threads* to be included in the corpus, but rather only to identify the *forums* that contained TM-related threads and required further investigation on a thread-by-thread basis. Both ProZ.com and TranslatorsCafé.com provide complex and simple searching of their discussion boards and forums, a necessary feature to perform a keyword search.

2.4.2 Creation of word list

Keywords were identified by creating a word list using the Word List function of WordSmith Tools.⁴⁵ A word list consists of an inventory and count of all single graphical words within a text or texts.⁴⁶

2.4.3 Selection of keywords

Over 2,000 unique graphical words were found in the sample corpus. This list was then lemmatized.⁴⁷ All words considered to be non-TM-specific and brand names were excluded from the list (e.g. file, document, Trados, Word). (See Appendix D.) The five most frequent, non-brand-specific, TM-related terms were chosen: *CAT*, *TMX*, *TM*, *memory/memories* and *segment/segments*.⁴⁸ I decided that if a forum contained occurrences of at least three of the keywords, it would be investigated on a thread-by-thread basis in order to extract all TM-related threads.

⁴⁵ WordSmith Tools is a corpus analysis tool developed by Mike Scott and distributed by Oxford University Press. For more information, see the WordSmith Tools website: <http://www.lexically.net/wordsmith/>.

⁴⁶ A graphical word is a string of characters delimited by spaces or punctuation. For example, “translation” is a graphical word, but “translation memory” is two graphical words.

⁴⁷ Lemmatization in WordSmith Tools is the process of combining word forms of the same lexical unit into one entry. For example, “memory” would be grouped with “memories”, and the number of occurrences of the two separate word forms would be combined to present a clearer indication of the frequency of the lexical unit in the text or texts.

⁴⁸ The inclusion of TMX is justifiable as it is entirely relevant to the use of TM systems and is not a proprietary format, but an open standard. The majority of TM systems on the market are TMX-compliant. See section 1.4.2.4.

2.4.4 *A priori* exclusions

Before performing a board-wide search, certain forums were excluded from the list of possible forums. Due to the specific topics of these forums, it was extremely unlikely that they would contain any TM-related threads, and thus it was expected that these forums would not offer a return on the investment of time necessary to perform a keyword search or in-depth investigation. For example, all language-specific forums were excluded *a priori* due to the language restriction (English only). Forums relating specifically to the use of ProZ.com and TranslatorsCafé.com were also excluded. In addition to these, any forums dedicated to highly specific topics that were not relevant to TM usage (e.g. Regional ProZ.com Conference forums, off-topic forums, terminology forums) were eliminated from the list. (For a complete and detailed listing of the forum search exclusions and results, see Appendix E.)

2.4.5 ProZ.com forum search

The discussion forums on ProZ.com can be searched easily and with flexible restrictions (e.g. by date, by forum, by author). For my purposes, I performed a search for each of the five keywords in the forums individually, searching both the posting title and posting content with no date restriction. If at least three keywords appeared anywhere in the forum, the forum was added to the list for further investigation.

The ProZ.com search feature does not accept a single search string that is fewer than three characters. This was problematic for the keyword *TM*; therefore, only four of the five keywords were searched in the ProZ.com forums. This character restriction posed a problem for three forums, “Linguistic Diversity”, “StudentZ Forum”, and “Teaching and Learning Languages.” All other forums already contained at least three keywords or fewer than two keywords. In the three instances mentioned above, an extra step was taken to evaluate the possibility of the forum containing useful threads. If less than 10% of the first 50 threads in the forum were TM-related, the forum would be disqualified from further investigation. In each of the three cases, none of the first 50 threads were TM-related, and the forums were excluded from the list for further investigation.

The keyword search eliminated five forums (the three mentioned above, “Translation Agencies”, and “Right-to-left Language Technical Forum”), leaving 28 forums on the list for further investigation. (For a complete and detailed listing of the forum search exclusions and results, see Appendix E.)

2.4.6 TranslatorsCafé.com forum search

The TranslatorsCafé.com search features are similar to those of Proz.com, and the same search restrictions used for the ProZ.com search were implemented for the TranslatorsCafé.com search. The TranslatorsCafé.com search engine did allow for the use of *TM* as a search string. The forum “Subtitling” contained only two keywords, and, as a result, was not included in the list. The search resulted in the addition of 16 forums to the list of forums to be investigated. (For a complete and detailed listing of the forum search exclusions and results, see Appendix E.)

2.4.7 *A posteriori* exclusions

I remained unconvinced of the usefulness of two extremely active and large forums, the “Viewpoints, Questions, Comments, Advice” forum on TranslatorsCafé.com and the “Translator Resources” forum on ProZ.com. Due to the sheer volume of threads in each of these forums (“Viewpoints, Questions, Comments, Advice” had 1365 threads on March 2, 2009 and “Translator Resources” had 3109 threads on March 6, 2009), I needed assurance of the return on the investment of time that a thread-by-thread investigation of these forums would require. An additional preliminary probe into their relevance to TM usage was made. The first 50 threads from each of the forums were analyzed, and, if less than 10% of the threads discussed the use of TM systems, the forum would not to be investigated further. The evaluation of “Viewpoints, Questions, Comments, Advice” revealed that only four of the first 50 results, or 8%, were relevant; therefore, the forum was eliminated from the list. The “Translator Resources” forum analysis, however, provided more promising results. Nine of the first fifty threads were TM-relevant (18%), and consequently, this forum retained its place on the list of forums to be investigated.

A “DTP – Desktop Publishing” forum on ProZ.com was created between the time of the keyword search and thread collection; therefore, as it had not undergone the keyword screening, it was eliminated as a possibility for further investigation.

2.4.8 Forum search summary

Out of a possible 129 forums on ProZ.com, 96 were eliminated *a priori*. An additional five forums were excluded on the basis of keyword search results. A total of 28 forums remained on the list to be analyzed on a thread-by-thread basis.

Out of a possible 50 forums on TranslatorsCafé.com, 33 were eliminated *a priori*. An additional two forums were eliminated based on keyword search results and a supplementary *a posteriori* probe for relevancy. A total of 15 forums remained on the list to be investigated on a thread-by-thread basis.

	ProZ.com	TranslatorsCafé.com
Total # of Forums	129	50
<i>A Priori</i> Exclusions	96	33
Keyword Search Exclusions	5	1
<i>A Posteriori</i> Exclusions	0	1
Total # of Forums Remaining	28	15

Table 10 - Forum Selection Results

At this point, the number of forums and volume of threads on ProZ.com was extremely daunting, and I had to consider my time restrictions. I reevaluated my decision to include both ProZ.com and TranslatorsCafé.com as primary resources and decided to retain the data from ProZ.com, as it seemed counterintuitive and unjustifiable to exclude ProZ.com on the basis that it contained *too much* potentially valuable data.

After identifying the forums to be investigated, a method for selecting TM-relevant threads was needed. This method is discussed in the next section.

2.5 Thread selection

The next step in the process of creating the corpus involved identifying the TM-related threads within the selected forums. Due to the number of forums and threads being considered, the date restriction used for the sample corpus was adjusted for the main corpus to obtain a more reasonable amount of data. For the creation of the main

corpus, only threads containing posts made on or after January 1, 2007 would be included.

The first step was to eliminate all the threads that had not been posted on since December 31, 2006. After excluding these threads, those remaining were examined, one by one, and the TM-related threads were identified and listed. This list included the forum title, thread title, posting count (i.e. the number of messages in the thread), and the date of the last posting. A total of 529 threads were gathered from ProZ.com, and 257 threads were gathered from TranslatorsCafé.com.

Due to the time involved in collecting and analyzing the data, all of these threads could not be included in the corpus. In order to gather the target number of posts (between 389 and 518 – see section 2.3) from each of the two discussion boards, a method of random selection was developed. The threads were sorted by posting date and then alphabetically by thread title. Every n^{th} thread was chosen, where n was the lowest possible number that yielded a posting total within the target range. The variable, n , was chosen to be the lowest possible number in order to obtain the greatest variety of threads and subsequently topics, as most threads are topic-specific. The value of n was twelve and three for ProZ.com and TranslatorsCafé.com, respectively. After completing the thread selection, the next steps were to finalize the corpus structure, create the corpus, and classify the postings. These steps are discussed in detail in the next sections.

2.6 Collection of postings

The final sample contained 937 postings, including both discarded and classifiable postings. A total of 358 classifiable postings were taken from ProZ.com and 360 from TranslatorsCafé.com. Both of these figures fell within the target range (between 300 and 400 classifiable postings).

Forty-four threads were taken from ProZ.com and 85 threads were taken from TranslatorsCafé.com. It is interesting to note that in order to reach the target range, just over half the number of threads were needed from ProZ.com in comparison to TranslatorsCafé.com, due to the higher number of postings per thread on ProZ.com.

Of the total 937 postings, 219 contained no classifiable information and were discarded. The final rate of discarded postings was 23%; this rate is identical to the

original calculation based on the sample corpus. A summary of these statistics is provided in Table 11.

	ProZ.com	TranslatorsCafé.com	Combined Total
Threads	44	85	129
Total Postings (including discarded postings)	440	497	937
Classifiable Postings	358	360	718
Percentage of Corpus – Classifiable Postings	49.86%	50.14%	100%
Discarded Postings	82	137	219
Percentage of Discarded Postings	19%	28%	23%

Table 11 - Summary of Corpus Statistics

2.7 Corpus structure

After creating the sample corpus and processing the data therein, I had a firm idea of how the main corpus should be structured. Before I began creating the main corpus, I established how the data would be organized and structured. The information in the corpus can be classified as either administrative or content-related.

2.7.1 Administrative information

Administrative information about each posting was included in the corpus. This information included the name of the forum in which the thread was found, the title of the thread, the title of the posting, the posting date, the posting author, the authors' country and languages (for posts from TranslatorsCafé.com) or translation language direction (for posts from ProZ.com), the date accessed, and the URL. In order to identify and easily refer to the threads and postings, each posting was given a unique identification number (e.g. TC28-04), including the site on which it was posted (P for ProZ.com or TC for TranslatorsCafé.com), the thread number (the number before the hyphen), and the posting number (the number after the hyphen); therefore, *TC28-04* refers to the fourth posting in the twenty-eighth thread collected from TranslatorsCafé.com.⁴⁹

⁴⁹ TranslatorsCafé.com labels each posting with its own unique number. These numbers were recorded in the corpus; however, I will be using my personally assigned numbers when referring to postings from TranslatorsCafé.com.

All of this information needed to be recorded, in the eventuality that I would need to refer to the original posting or webpage. Furthermore, as I would be analyzing and commenting on information about the authors, their locations, and working/native languages, this information needed to be included. For an example of how the administrative data was tracked in the corpus, see Appendix F.

2.7.2 Content-related information

Content-related information consisted of the body of the message, or the posting itself, as well as its categorization according to the classification system. (See section 2.7.3 for a detailed description of the creation and implementation of the classification system.) The thread title and posting title were not considered to be content-related.

Quoted articles, websites, books, and other references to sources outside of the discussion board were not included in the postings, as they do not contain the opinion or viewpoint of the author, but of a third party.

Due to the responsive nature of threads, posts often contain quotations from previous posts, either from within the thread or from other threads. Extra caution was taken in these instances to avoid duplication when classifying the content. Only the original content of each posting was considered for classification. See Appendix G for an example of the organization of the content-related information.

2.7.3 Classification system

The classification system for the main corpus was based on the preliminary system developed for use with the sample corpus and was divided into three main sections: topic-based categories, brand-name categories, and operating-system categories. The categories were organized in a hierarchical structure, with a main topic or brand category divided into topic or brand subcategories. If a posting referred to a topic, brand name, and/or operating system, an indicator was added to the corresponding column(s). The classification system was not definitive before I began creating the corpus, as it would have been impossible to predict all of the necessary categories. Therefore, categories were added, removed, and redefined throughout the process.

2.7.3.1 Topic-related categories

The main component of the classification system consisted of the topic-related categories. A given posting could be classified into one or several (sometimes 10 or more) of these topic subcategories; however, as threads and postings tend to deal with a somewhat restricted subject or topic, this number was generally not high. Table 12 lists the topic-specific categories used for classification.

TEnT Components		alignment (including segmentation)
		statistical document analysis (including word count)
		concordancing
		localization features
		machine translation
		quality assurance features
		project management features
		terminology management system
Purchasing or Acquiring a TM System		purchasing or acquiring a TM system (general)
	Situational Considerations	text characteristics and format
		networking (number of users)
		languages (interface, source and target)
	Technical Considerations	computer capabilities (processing speed, memory)
		operating system
		open-source software
		word-processor or text-editor integration
		voice recognition software
	Brand-specific Considerations	trial or demo versions
		price
		updates or new releases
Using a TM System	Training	training
	File Formats	source and target file formats (including formatting)
		legacy file formats
		files created by the TM system
		TM formats
		compatibility among TM systems
		vendor lock-in
		open standard formats (SRX, TMX)
	Matches	types of matches
		usefulness of matches
	Technical Issues	customizability of display and keyboard shortcuts
		technical problems (general)
		technical support
	Procedural and Textual Effects	speed
		quality
		modifications to the translation process

	Client Relations	brand or format requirements
		payment
		ownership
		buying/selling/trading TMs
Other		non-adoption attitudes
		vendor promotional materials
		other forums (Yahoo, ProZ.com, TranslatorsCafé.com)

Table 12 - Hierarchical Listing of Topic-Related Categories

2.7.3.2 Brand-name categories

The brand-name categories represent the specific tools discussed within the postings. The majority of these categories are brand names of TM systems or TEnTs; however, other tools which commonly interact with TM systems were also tracked (e.g. voice recognition software, terminology management systems, word processors). When a specific component of a TEnT was being discussed, categories for both the component and the TEnT brand were indicated. Note that all in-house TM systems were grouped together in one category. In some postings, the authors referred to other computer software formats containing the text to be translated (e.g. AutoCAD or InDesign files). Categories for these applications were not created, and, in these instances, the issue was tracked under a topic-specific category such as “source and target file formats”. Table 13 lists the brand-name categories used for classification.

TEnTs	Across
	Across CrossTank
	Anaphraseus
	AppleTrans
	Déjà Vu
	Fortis
	Fusion
	Heartsome
	IBM Translation Manager
	Idiom Worldserver
	In-house TM System
	Lingotek
	Logoport
	MetaTaxis
	MemoQ
	MultiTrans
	Open Language Tools
	OmegaT

	Similis
	SDL Trados
	MultiTerm (terminology management system)
	Passolo (localization tool)
	SDLX (TM system)
	TagEditor (TM system)
	Translator's Workbench (TM system)
	WinAlign (bibtex aligner)
	Swordfish
	Transit
	TranslationOffice 3000
	Wordfast
Localization Tools	Alchemy Catalyst
	CatsCradle
	ForeignDesk
	Lingobit
	RC WinTrans
	Sisulizer
Voice Recognition Software	DragonNaturally Speaking
OCR Software	Fine Reader
File Searching Software	Copernic Desktop Search
Word Processors, Text Editors, and Office Suites	Generic Text/XML Editor
	Macintosh iWorks
	Microsoft Office
	OpenOffice
Add-ons, Plug-ins, and other TM Tools (for conversion, tag processing, compliance checking)	+Tools
	Comparator
	DWG Translator
	Olifant TM Editor
	PlusToyz
	TMXValidator
	Tortoise Tagger
	TTXPress
	Xbench

Table 13 - Hierarchical Listing of Brand-Name Categories

2.7.3.3 Operating-system categories

The operating-system categories were used to track the discussion of specific OSs. (See section 1.3.2.1.2 for a description of an OS and its relation to TM usage). This group of four categories allowed for the indication of discussions of particular OSs (Linux, Macintosh, Windows, and Other). The discussions relating to specific OSs were also tracked by a generic OS topic-related category (in the same fashion as tool

components were tracked individually and under the generic brand-name category). The topic-related category was necessary to trace discussions which mentioned OSs but did not refer to a specific system.

2.8 Concluding remarks

In this chapter, I have discussed in detail the steps and processes involved in selecting the type of primary source to be used, the individual discussion boards to be investigated, and the threads to be included in the corpus. I have also described the design and creation of the corpus and classification system. The analysis of the opinions, questions, comments, and viewpoints contained therein are discussed in the next chapter.

Chapter 3 – Corpus Analysis

Using the methodology outlined in Chapter 2, I was able to collect and classify a large number of English-language postings relating to TM usage from the discussion boards on ProZ.com and TranslatorsCafé.com. The analysis of these postings has provided interesting and informative results which are presented in this chapter, highlighting numerous instances where users are divided on TM-related issues, indicating strengths and weaknesses of particular brands, and revealing gaps in and incongruities with current TM-related research. Section 3.1 provides statistics relating to the contributors and number of postings, and discussion surrounding the most and least commonly addressed topics, brands, and operating systems. Section 3.2 presents an analysis of the postings, organized thematically based on the structure of the literature review in sections 1.3 and 1.4.

3.1 Corpus statistics

The final corpus contained 718 postings from a large number of unique authors, working in different languages, from around the world.⁵⁰ This variety helped to ensure that a wide range of opinions was reflected in the data. A breakdown of these statistics is provided in Table 14.

⁵⁰ Note that in this chapter, the terms *authors*, *translators*, and *users* all refer to the authors of postings included in the corpus.

	ProZ.com (P)	TranslatorsCafé.com (TC)	Total
Threads	44	85	129
Postings	358	360	718
Unique Authors ⁵¹	160	136	296 ⁵²
Countries ⁵³	43	39	53 ⁵⁴
Language or Language Direction (P) ⁵⁵	54	41	n/a ⁵⁷
Mother Tongue(s) (TC) ⁵⁶			

Table 14 – General Corpus Statistics

3.1.1 Topic-related category statistics

A total of 46 topic-related categories were used to classify the postings and a grand total of 1702 topic-related classifications were made.⁵⁸ The categories and related statistics are listed in Table 15. The percentage indicates the proportion of total postings which contained references to that particular topic. Table 15 does not include statistics related to discussions of other TEnT components.

⁵¹ ProZ.com and TranslatorsCafé.com members are required to indicate a username, which was used to identify the unique authors.

⁵² It seems as though there are seven instances of direct overlap between the two sites (i.e. authors with identical usernames posting on both sites); however, it cannot be assumed that these usernames belong to the same individuals. It is possible the corpus contains further instances of overlap if authors were using a different username for each site.

⁵³ ProZ.com and TranslatorsCafé.com members are not required to indicate their country of residence; therefore, this information is unavailable for some authors.

⁵⁴ This number is not a sum of the ProZ.com and TranslatorsCafé.com totals, as there were authors from the same countries posting on both sites. This figure indicates the number of unique countries.

⁵⁵ ProZ.com members may indicate their language direction for translation; however, this field may be left blank or list only one language. This figure indicates the number of unique languages or language directions.

⁵⁶ TranslatorsCafé.com members may indicate multiple languages as mother tongues. This figure indicates the number of unique mother tongues or unique combinations of mother tongues.

⁵⁷ This total cannot be calculated as the sites display different language information (see above).

⁵⁸ All 46 categories are not listed in the table, as some categories were further divided in order to specifically track certain aspects of a given issue (e.g. the “source and target file formats” category is actually comprised of two separate categories: source and target file formats and text formatting).

Topic	Number of References	Percentage of Total # of Postings Referencing Topic
<i>Purchasing or Acquiring a TM System</i>	-	-
purchasing or acquiring a TM system (general)	221	30.8%
<i>Situational Considerations</i>	-	-
text characteristics and format	36	5.0%
number of users	15	2.1%
languages (interface, source and target)	51	7.1%
<i>Technical Considerations</i>	-	-
computer capabilities (processing speed, memory)	14	1.9%
operating system	41	5.7%
open-source software	18	2.5%
word-processor or text-editor integration	30	4.2%
voice recognition software	8	1.1%
<i>Brand-Specific Considerations</i>	-	-
trial or demo versions	51	7.1%
price	87	12.1%
updates or new releases	49	6.8%
<i>Using a TM System</i>	-	-
<i>Training</i>	-	-
training	78	10.9%
<i>File Formats</i>	-	-
source and target file formats	168	23.4%
files created by the TM system	96	13.4%
TM formats	76	10.6%
compatibility among TM systems	46	6.4%
vendor lock-in	10	1.4%
open standards – SRX	2	0.3%
open standards – TMX	39	5.4%
<i>Matches</i>	-	-
types of matches	12	1.7%
usefulness of matches	8	1.1%
<i>Technical Issues</i>	-	-
customizability of display and keyboard shortcuts	22	3.1%
technical problems (general)	156	21.7%
technical support	39	5.4%
<i>Procedural and Textual Effects of TM Usage</i>	-	-
speed	21	2.9%
quality	11	1.5%
modifications to the translation process	7	1.0%
<i>Client Relations</i>	-	-
brand or format requirements	138	19.2%
payment	42	5.8%
ownership	12	1.7%

Topic	Number of References	Percentage of Total # of Postings Referencing Topic
buying/selling/trading TMs	6	0.8%
<i>Other</i>	-	-
non-adoption attitudes	44	6.1%
vendor promotional materials	32	4.5%
other forums	16	2.2%
Grand Total	1702⁵⁹	

Table 15 - Topic-Related Category Statistics

Table 16 lists the topic-related categories sorted by frequency. The most commonly discussed topics were (in decreasing order of frequency and excluding the general categories of “purchasing or acquiring a TM system” and “technical problems”): source and target file formats (23.4%), client brand or format requirements (19.2%), files created by the TM system (13.4%), price (12.1%), and training (10.9%). Topics discussed in the literature review but not mentioned in the corpus were: the language of the system’s interface, methods of TM creation (interactive versus post-translation alignment), storage and retrieval units (SB versus CSB), methods of TM use (interactive mode versus pretranslation), and the format of legacy files. The lack of representation of these topics is significant in the evaluation of the corpus, and possible explanations for this are discussed in sections 4.1, 4.2, and 4.3.

⁵⁹ Note that this number represents the total number of classifications; one posting could be classified under numerous categories.

P/A, U or O ⁶⁰	Topic	Number of References	Percentage of Total # of Postings Referencing Topic
U	source and target file formats	168	23.4%
U	brand or format requirements	138	19.2%
U	files created by the TM system	96	13.4%
P/A	price	87	12.1%
U	training	78	10.9%
U	TM formats	76	10.6%
P/A	languages (interface, source and target)	51	7.1%
P/A	trial or demo versions	51	7.1%
P/A	updates or new releases	49	6.8%
U	compatibility among TM systems	46	6.4%
O	non-adoption attitudes	44	6.1%
U	payment	42	5.8%
P/A	operating system	41	5.7%
U	open standards – TMX	39	5.4%
U	technical support	39	5.4%
P/A	text characteristics and format	36	5.0%
O	vendor promotional materials	32	4.5%
P/A	word-processor or text-editor integration	30	4.2%
U	customizability (display, keyboard shortcuts)	22	3.1%
U	speed	21	2.9%
P/A	open-source software	18	2.5%
O	other forums	16	2.2%
P/A	number of users	15	2.1%
P/A	processing speed and memory	14	1.9%
U	types of matches	12	1.7%
U	ownership	12	1.7%
U	quality	11	1.5%
U	vendor lock-in	10	1.4%
P/A	voice recognition software	8	1.1%
U	usefulness of matches	8	1.1%
U	modifications to the translation process	7	1.0%
U	buying/selling/trading TMs	6	0.8%
U	open standards – SRX	2	0.3%

Table 16 - Topic-Related Categories by Frequency

⁶⁰ P/A – Purchasing or Acquiring a TM System

U – Using a TM System

O – Other

3.1.2 Brand-name category statistics

A total of 45 brand-name categories were used to classify the postings, and a grand total of 1152 brand-name classifications were made (these figures exclude sub-categories within brands and only include the generic brand-name category.) The categories and related statistics are listed in Table 17. The percentage indicates the proportion of total postings which contained references to that particular brand.

Tool Type	Brand Name	# of References	% of Total # of Postings with Reference
Translation Environments	Across	22	3.1%
	Anaphraseus	3	0.4%
	AppleTrans	1	0.1%
	Déjà Vu	59	8.2%
	Fortis	9	1.3%
	Fusion	1	0.1%
	Heartsome	16	2.2%
	IBM Translation Manager	1	0.1%
	Idiom Worldserver	12	1.7%
	In-house TM System	4	0.6%
	Lingotek	1	0.1%
	Logoport	1	0.1%
	MemoQ	34	4.7%
	MetaTaxis	37	5.2%
	MultiTrans	2	0.3%
	Open Language Tools	1	0.1%
	OmegaT	54	7.5%
	Similis	6	0.8%
	SDL Trados	452	63.0%
	Swordfish	3	0.4%
	Transit	17	2.4%
	TranslationOffice 3000	1	0.1%
	Wordfast	138	19.2%
Localization Tools	Alchemy Catalyst	5	0.7%
	CatsCradle	2	0.3%
	ForeignDesk	1	0.1%
	Lingobit	1	0.1%
	RC WinTrans	4	0.6%
	Sisulizer	14	1.9%
VRS	DragonNaturally Speaking	10	1.4%
OCR Software	Fine Reader	4	0.6%
File Searching Software	Copernic Desktop Search	2	0.3%

Word Processors, Text Editors, and Office Suites	Generic Text/XML Editor	20	2.8%
	Macintosh iWorks	2	0.3%
	Microsoft Office	168	23.4%
	OpenOffice	11	1.5%
Add-ons, Plug-ins, and other TM Tools (for conversion, tag processing, compliance checking)	+Tools	9	1.3%
	Comparator	4	0.6%
	DWG Translator	2	0.3%
	Olifant TM Editor	3	0.4%
	PlusToyz	4	0.6%
	TMXValidator	1	0.1%
	Tortoise Tagger	2	0.3%
	TTXPress	5	0.7%
	Xbench	3	0.4%
Grand Total		1152	

Table 17 - Brand-Name Category Statistics

Table 18 lists the distribution of brand-name references based on tool type. Specific TM brand names were mentioned in 76% of all postings, and brand names of office suites, word processors, and text editors were mentioned in 17.4% of postings.

Type of Software	Total Number of References	Percentage of Brand References for Type of Software
Translation Environments TM Systems	875	76.0%
Localization Tools	27	2.3%
Voice Recognition Software	10	0.9%
OCR Software	4	0.3%
File Searching Software	2	0.2%
Word Processors, Text Editors, and Office Suites	201	17.4%
Add-ons and Plug-ins	33	2.9%

Table 18 - Brand References Based on Tool Type

Table 19 lists the five most commonly discussed TEnT brands. Although no solid conclusions regarding tool choice and usage can be made, it is interesting to compare these results to those of the LISA 2004 (Lommel 2004) and Lagoudaki (2006) surveys (also listed in Table 19), noting the continued prevalence of SDL Trados and Déjà Vu, but also the discussion of lower-cost and free TEnTs, such as Wordfast and OmegaT, and the appearance of a new tool on the list, MetaTaxis. (Note that although MetaTaxis does not appear on the LISA or Lagoudaki lists, it was available at the time of surveys; it was released in 2002.)

Brand	# of References in Corpus	Placement in Corpus Top Five	Placement in LISA 2004 (Lommel 2004)	Placement in Lagoudaki (2006)
SDL Trados	452	1	1	1
Wordfast	138	2	6	2
Déjà Vu	59	3	3	4
OmegaT	54	4	10	8
MetaTaxis	37	5	-	-

Table 19 - Comparison of Corpus, LISA 2004, and Lagoudaki Top Tools

3.1.3 Operating-system category statistics

Macintosh operating systems were the most discussed in the corpus (40 references), closely followed by Windows (37 references). This may be due to the number of Macintosh users investigating TEnTs that are compatible with their OS (see section 3.2.1.2.1.2 for further discussion). Table 20 lists the statistics relating to the operating-system categories.

Operating System	# of References	% of Total # of Postings with Reference
Linux	11	1.5%
Macintosh	40	5.6%
Windows	37	5.2%
Other	4	0.6%
Grand Total	92	

Table 20 - Operating-System Category Statistics

3.2 Analysis and summary of corpus content

The content of the postings is discussed thematically based on the classification system and the structure of sections 1.3 and 1.4. Each subsection provides a brief overview of the corpus statistics and a synthesis of the content in the corpus using verbatim quotations from the postings. The corpus findings are then compared and contrasted to the literature review in Chapter 1. My opinions and interpretations are discussed in the conclusion. (Appendix A contains a listing of the issues addressed in Chapters 1 and 3 and the conclusion and provides cross-references to these discussions within the thesis.)

It is necessary to note that many of the authors are not native speakers of English, and the language used in discussion forums is often informal; thus, the quotations contain grammatical and spelling errors. These errors have been left intact, without the use of

[sic] in order to minimize disruptions in the flow of the postings. Additionally, to facilitate the identification of direct quotations, these are displayed in *Arial Narrow font*. Postings are cited using personally assigned identifiers (see section 2.7.1). Appendix H contains the complete reference details for cited postings.

3.2.1 Purchasing or acquiring a TEnT

A total of 221 postings (30.8% of all postings) related to purchasing or acquiring a TEnT. The majority of these discussions began with a new or potential user asking advice about which TM system he or she should purchase and whether it would be a worthwhile investment. While some did respond to these enquiries with short replies such as “*De javu is the best. wordfast is the cheapest. Trados is the most popular. SDLX is the easiest*” (TC36-02), it was reassuring to observe that most participants recognized that “*there is no such thing as 'the best' CAT tool. The 'best' one is the one which suits *you* best*” (TC36-03). The decision to purchase a TM system should not be taken lightly, and, as seen through the analysis of the corpus, the members of ProZ.com and TranslatorsCafé.com seemed to enjoy sharing their opinions and advice with their peers. The following section highlights discussions relating to the factors considered when purchasing or acquiring a TM system.

3.2.1.1 Situational considerations

3.2.1.1.1 Text characteristics and format

Authors discussing the usefulness of CAT tools (particularly TM systems) and text characteristics (e.g. amount of repetition, text author, subject field, text type) seemed to agree that these tools are most beneficial when translating repetitive texts. Out of a total of 36 postings related to text characteristics and usefulness of TM systems, the majority, 25 postings, indicated that repetition increases usefulness. Furthermore, in some of these postings, repetition was also linked with subject field. For example, one author wrote, “*In your areas of expertise - Zoology & Botany, Environment & Ecology, Agriculture, General Science etc. - texts with highly repetitive segments (sentences) are a rarity. [...] Hence, any CAT tool would be of little use, if any at all*” (P20-02). Another user made an

association between text type (and repetition) and usefulness: “*It [Trados] has worked for me so far, as I mostly translate technical manuals with many repetitions*” (TC60-06).

In addition, opinions regarding usefulness were also based on the source of the work; one author wrote, “*If you do a lot of work for agencies with big and/or ongoing projects allocated over several translators, then it [Trados] is probably worth it. If you have a combination of direct customers and agencies that do not require it, then you don't [need it]*” (P18-02). In this case, the author’s view of TM usage is that the choice may be determined by the requirements of the client or agency or scope of the project. Issues relating to client or agency requirements are discussed in detail in section 3.2.2.6.1.

While the majority of postings relating to text type, subject field, and repetition showed a perceived relationship with usefulness, opinions diverged when discussing other texts that may be translated using TM systems. Some users believed that TM systems (and other TEnT components) are beneficial regardless of repetition, and used them for all translation work. One such user posted:

I do even non-repetitive texts in my CAT tool (DejaVuX, a.k.a. DVX), and I find it a good way to benefit from my own terminology [...] I find that this helps me to be consistent, and it also avoids the experience of looking up key concepts several times during a big job. And if I come across words that I have translated before but not recorded as terminology items, I can find them by doing a concordance search in the sentences database (= translation memory). (P20-03)

In fact, other postings revealed opinions indicating that TM systems are a requirement for translators, and, as this author wrote, are necessary “*unless you're translating only faxes, PDF documents and hardcover books*” (P18-05).

One thread in particular addressed the possibility of using TM systems to translate fiction. Discussion began in response to a poll asking, “*Should using CATs for translating fiction be banned?*” (TC48). The discussion surrounding the poll revealed that five out of the six participants felt that the use of TM systems could be beneficial when translating texts that were not repetitive. Authors wrote that some may find it “*practical to translate with segmented texts while being still free to access TM's and glossaries*” (TC48-01), users may “*prefer to work in its [the TM system's] translation environment*” (TC48-02), the tool may “*help with context-specific words, terms, abbreviations*” (TC48-05), a TM system is useful to “*keep the format of the original or to check a certain phrase in the memory*” (TC48-03), or that having

“everything on the screen in front of me is better for my concentration [...] and ensures that I do not skip sentences when looking from the source to the target texts” (TC48-08).

The usefulness of TM systems for repetitive texts is clearly established in the literature (see section 1.3.1.1). This apparent relationship between text type (and repetition) and TM usage was also evident in the findings of Lagoudaki’s 2006 survey. Lagoudaki notes, “It was confirmed that those who specialise in technical texts are more likely to use TM tools [...] unsurprisingly, the TM usage rate drops for those who translate general texts (or have no specialisation) and for those who translate literature” (2006, 16). The use of TM systems for non-repetitive texts is perhaps a more recent development, and although not noted in Lagoudaki’s survey, this observation was made by García in 2006. García noticed an emerging “body of thought” on Lantra-L surrounding the use of TM systems for non-repetitive texts, made possible through the use of tools such as Déjà Vu’s Assemble (example-based machine translation) and sub-segment matching (2006, 102). This “body of thought” is also present on the discussion boards examined; authors did, in fact, indicate that the use of TM systems for translating non-repetitive texts can also prove useful. While the subject field and amount of repetition are largely motivating when deciding to use a TM system, perhaps these are not the only factors which affect the benefits of TM usage.

It is interesting to note the lack of discussion in the corpus relating to hard-copy documents. Message authors rarely referred to the format of the texts they were translating (i.e. hard copy or electronic), and, in fact, only two authors indicated that they were receiving work in the form of faxes. The corpus contains no discussions relating to the use of voice recognition or OCR software for conversion of hard-copy documents to electronic format. In contrast, two studies have shown that the use of hard-copy source documents is a common deterrent to TM adoption: Lagoudaki noted that the use of hard-copy documents was the number one reason given for non-adoption of TM systems (2006, 20); Lommel observed that “surprisingly, a number of large organizations” indicated use of hard-copy documents as a reason for non-adoption (2004, 6). Furthermore, the issue of conversion to electronic format is also addressed in literature such as Bowker (2002) and L’Homme (2008). Possible reasons for lack of discussion in

the corpus about hard-copy documents and their conversion are examined in sections 4.1, 4.2, and 4.3.

3.2.1.1.2 Number of users

The corpus contains very few postings (15) discussing the networking capabilities of TM systems. The exceptions include a discussion about which tool might be best for a collaborative project (among numerous users accessing the same files) (TC07) and a discussion about installing Trados on two home computers concurrently, but not using the software on both computers simultaneously (TC51). As noted in section 1.3.1.2, the networking capabilities of a TM system are perhaps a more important consideration for agencies or organizations, rather than freelance translators; therefore, this lack of discussion may be due in part to the type of translators whose postings are included in the corpus. This observation is discussed further in sections 4.1, 4.2, and 4.3.

3.2.1.1.3 Languages

3.2.1.1.3.1 Language pairs

The discussions surrounding language and TM systems were mainly focused on language-specific difficulties, such as problems with inserting accented or special characters or storing these characters. Some users may not be using Unicode (a standard for storing characters), which can cause character-related issues when saving and opening TMs. For example, one author was having problems inserting uppercase accented Romanian characters and, rather than using Unicode, was inserting lowercase characters and applying an uppercase style (TC24). Another experienced user pointed out that one should be cautious with this method, as “*Word does not properly stores those accented characters [...] If someone somehow resets the characters styles of the text for instance (like by pressing CTRL+SPACE), your precious layout will go down the drain and you'll be left with improper characters*” (TC24-04). However, with the use of Unicode, the majority of problems can be overcome, as indicated by one author: “*In Déjà Vu everything is done in Unicode and thus there is no limitations in terms of languages, even when it comes to complex languages such as East Asian double-byte languages, or right-to-left languages such as Arabic and Hebrew*” (TC07-02).

Although the need for improvement of TM systems to handle right-to-left languages and double-byte characters was identified in the literature (see section 1.3.1.3.1), little discussion of these issues was found in the corpus. The threads relating to right-to-left languages did not directly address a lack of support: one author was seeking help with technical language-related problems (TC17) and another was seeking a TM system for a collaborative project (TC07). (Fifteen authors indicated either Arabic or Hebrew as their mother tongue or one of their working languages (13 Arabic, two Hebrew).)⁶¹

Only one thread addressed an issue of insufficient language support: the translator was working with Bengali, but found that SDLX did not support that language (TC27-08). Other language-related threads discussed the limitation of five working languages imposed in SDL Trados (TC11-02, P18-08), inversion of source and target languages of TMs (P27-02), and voice recognition/TM software conflicts when inserting accented or special characters (P17-01).

The discussions related to language pairs presented somewhat isolated cases that, for the most part, were resolved with the help of other forum members, perhaps indicating that either translators are having no widespread difficulties concerning language support or availability, or that these issues are not among the main concerns of TM users. In the LISA 2004 survey, Lommel noted that only 5% of respondents who were not using TMs for all of their languages (i.e. working languages of a translator, agency, or organization) indicated a lack of language support as the contributing factor for non-adoption (2004, 7). Lommel attributes this to “recent developments within TM tools and within external applications with which they interface, particularly with regard to Unicode support” (Ibid.). Furthermore, the issues associated with right-to-left languages and double-byte characters noted by Bowker (2002, 120) may no longer be problematic, as Quah writes that “virtually all translation memory systems are language-independent and support international character sets that represent many, if not all, alphabets and scripts digitally” (2006, 94). The shift in attention from 2002 (Bowker) to 2004 (Lommel) to 2006 (Quah) perhaps indicates that previous issues have been

⁶¹ It is interesting to note that ProZ.com does have a specific forum dedicated to right-to-left languages; however, at the time the corpus was created the forum was empty.

addressed, thus affecting the amount of discussion related to language support. Factors contributing to a possible under-representation of these issues in the corpus are further discussed in sections 4.1, 4.2, and 4.3.

3.2.1.1.3.2 Language of the interface

Despite the relationship between learning curve and the language of the interface as noted in the literature review (Bowker, McBride, and Marshman 2008, 31), no postings in the corpus discussed the availability or benefits and drawbacks of the language of the interface. This may be due in part to the methodology implemented and the type of authors whose postings were included in the corpus; these issues are further addressed in sections 4.1, 4.2, and 4.3.

3.2.1.2 Technical considerations

3.2.1.2.1 Computing requirements

3.2.1.2.1.1 Processing speed and memory

Fourteen postings related to the processing and storage requirements needed to properly operate a TEnT, 11 of which discussed the memory and processing needs of the software and three of which discussed the effects of using large TMs. For example, one author noted that “*the program [Déjà Vu] is even more of a memory hog than Trados*” (P18-15), another noticed that the use of Across tended to “*slow down the computer*” (TC13-05), and another wrote that “*Wordfast doesn't have those speed issues [that Trados does]*” (TC54-05). One thread (P38) specifically addressed the potential problems of running SDL Trados on a Netbook (a small laptop computer with a lower processing speed and less memory than a standard laptop).

The size of the TM was identified as problematic in three postings; the users' systems were unable to open and use the TM as it required too much memory. (These large TMs could be anywhere from hundreds of megabytes to a gigabyte.) As discussed in section 1.4.3.2, some users may prefer one large TM versus smaller, more specific TMs, and this can lead to processing issues (see section 3.2.2.3.2 in the analysis for further discussion of division of TMs).

3.2.1.2.1.2 Operating system

The Macintosh OS was the most commonly mentioned in the corpus (40 references), followed by Windows (37 references), and Linux (11 references). The most common issues relating to tool choice and OS were addressed by non-Windows users, that is users of Macintosh or Linux systems. In fact, 17% of all postings relating to purchasing a TM system were made by users specifically researching non-Windows tools. In most cases, members were quick to list Macintosh- or Linux-compatible tools (e.g. OmegaT, Heartsome, Wordfast). One posting generally described the end-user base for Macintosh or Linux systems:

Having worked for several translation tools providers (and attended Mac tradeshow for that purpose), I have analyzed the market segments for PC and Mac based translation tools over the past decade. All of the survey results showed that the Mac-based professional translator community was actually quite limited, and did not justify creating and maintaining Mac-specific products [...] [the number of] end-users worldwide is nowhere the user base necessary for a dedicated Mac product. [...] Such a number hardly justifies creating a Mac-designed product, and maintaining a regular product life cycle of builds, testing, support, patches, etc for Mac users. (TC83-04)

Although this may describe the past and even current situation of Macintosh and Linux TM systems, the increasing market share of Linux and Macintosh systems may reopen this door for further investigation and development.

Some users who were unable to run the system of their choice (which meets their needs and their clients' requirements) on Macintosh or Linux systems were installing software such as Parallels Desktop, a program which allows a user to operate a Windows OS on a Macintosh or Linux computer. By creating this type of "virtual machine", a user does not have to sacrifice his or her preferred OS for the sake of the TEnT or TM system. Although this may appear to be a viable solution, the decision to use Macintosh or Linux systems is often based on a rejection of Windows; TM users who have rejected Windows tend not be enthusiastic about operating within the Windows OS and may find this option to be counterintuitive. As one author pointed out, "*What's the fun of having a Mac and [using] Windows?*" (P26-08). Furthermore, the use of software such as Parallels Desktop may present keyboard issues such as shortcut conflicts or missing and/or different keys on the keyboard.

As described in section 1.3.2.1.2, the need for and number of Macintosh- and Linux- compatible tools is growing, and the availability of cross-platform tools could, as Bowker, McBride, and Marshman point out, potentially increase TM uptake (2008, 38). The availability (or unavailability) of tools for a user's preferred OS was identified as contributing to non-adoption in the findings of the LISA 2004 survey (Lommel 2004, 7). Although Lagoudaki's findings show that the majority of respondents used Windows (92% Windows, 4% Macintosh, 2% Linux, and 2% a combination of OSs) (2006, 13), the number of postings addressing the issue of OS and tool choice demonstrates that users are, in fact, investigating tools that run on an OS other than Windows. It is also interesting to note the appearance of OmegaT, Star Transit, and Wordfast, all cross-platform tools, in the top ten most commonly used TM systems of the LISA 2004 and Lagoudaki surveys (Lommel 2004, 12; Lagoudaki 2006, 24). While conclusions cannot be drawn as to whether or not their presence in the list is due to their OS compatibility, it does indicate that tools that are only available for Windows are not the only tools being used.

3.2.1.2.2 Open-source versus closed-source software

The corpus contains 18 references to the use of open-source tools, especially in discussions relating to cross-platform capabilities (perhaps due to the use of Linux, an open-source OS) and price (open-source software is usually available free of charge) (see sections 3.2.1.2.1.2 and 3.2.1.2.2 in the analysis). The most well-known open-source TM system is OmegaT, and authors often recommend it as a starting point for new users. In fact, OmegaT was the fourth most commonly discussed tool. This enthusiasm may be due to the fact that there is no financial risk involved (as the tool is fully functional and free, forever), or it may be due to OmegaT's cross-platform capabilities (as OmegaT is available for users of Windows, Macintosh, and Linux OSs). As the open-source movement gathers momentum and its principles and practices are promoted, perhaps translators will become more aware of the potential and quality of open-source translation tools.

3.2.1.2.3 *Word-processor or text-editor integration*

Discussions surrounding word-processor integration revealed a divide among TM users. Some enjoy word-processor or text-editor integration (e.g. integration with third party software such as Microsoft Office Word rather than incorporation of a proprietary text editor), while others prefer working directly within the system's proprietary text editor. In the 30 postings relating to TM systems and this type of integration, three authors indicated they prefer word-processor or text-editor integrated environments, six authors stated they prefer proprietary environments, and the remaining 21 posts showed no inclination towards either. One Déjà Vu user wrote, *"When you work with DVX you often don't need even to be familiar with the applications which file formats you use because every task within Déjà Vu is performed from and within its own user interface"* (TC45-01). Some may appreciate the simplicity of non-integrated environments: one Across user stated, *"It is all in one application and you don't have to start several windows"* (TC13-03), and some may find a TM system's own interface to be *"more pleasant and convenient"* than that of integrated third party software (TC05-02).

Others have had negative experiences with proprietary text editors: *"I was never comfortable with tools that required the use of their own environment, as I once got stuck when using one and had to copy and paste every sentence of the job to get it 'out'"* (TC05-03). While opinions may differ, it is important to recognize that a user's preference will depend on the situation and his or her requirements for word processing functions. As one author pointed out, *"Wordfast's advantage is that it runs in an environment familiar to virtually all translators. It's disadvantage is that it is limited by this same environment"* (P12-14).

Although the integration of TM systems with third party word-processors has been a common practice in the past, one author observed *"that the integration of CAT tools with Word is getting less and less. Trados seems to concentrate more on TagEditor and SDLX, Wordfast is about to come with a standalone version, Deja Vu has been standalone since more than 10 years"* (P29-02) and later iterated that this integration might be *"on its return"* (i.e. no longer as common) (P29-06).

As mentioned in the literature review in section 1.3.2.3, integration with a familiar program may shorten a user's learning curve; however, this pairing of software may necessitate additional training for a user to become comfortable with both applications.

Furthermore, the use of a system with an integrated word processor or text editor may limit file format support, create compatibility issues among brands, and increase the possibility of vendor lock-in. (See sections 3.2.2.2.1, 3.2.2.2.2, and 3.2.2.2.3 for further discussion of these issues.) Lagoudaki's findings indicated that 56% of respondents found Microsoft Word (i.e. an integrated word processor) to be the most practical environment for text processing, 22% preferred processing text in the TM system's own environment (i.e. using a proprietary text editor), 20% indicated that either of the two previous options were practical, and 3% indicated other various text editors or word processors to be the most practical (2006, 30). These respondents are somewhat divided on the issue, and this division is reflected in the discussions in the corpus related to integrated versus proprietary text editors or word processors.

3.2.1.2.4 Voice recognition software

Discussions relating to the use of voice recognition software (VRS) and TM systems were not extremely common in the corpus (a total of nine postings referencing VRS usage with TM systems). These authors primarily addressed compatibility and technical issues between VRS packages and TM systems. The strongest complaint related to processing speed, and, according to one user, a VRS package requires “*an enormous amount of memory, esp. if used in tandem with TM/translation memory [...] [and] is often overtaken by updates that need even more memory. [...] [These factors] compound the memory capacity problem and are even apt to bring down the whole PC operating system*” (P31-10). Furthermore, the interaction between the VRS and TM system can cause both conflict issues when giving commands and insertion problems when dealing with tagged files (P17-01, P07-03, TC18-02). While the benefits may be obvious (see section 1.3.2.4), users who wished to use VRS with their current TM system seemed to be aware of the potential issues and were investigating the compatibility of the specific applications before making their purchase.

3.2.1.2.5 TM creation, storage, and usage methods

Despite the attention paid to the different methods of TM creation (e.g. interactive creation or post-translation alignment), storage and retrieval units (SB and CSB

approaches), and usage methods (interactive or batch mode) in the literature review (see section 1.3.2.5), no discussions in the corpus pertained to the potential benefits and drawbacks of these methods. This may indicate a lack of awareness, or it may be a reflection of the methodology implemented and the types of users whose postings were included in the corpus; the possibility of these as contributing factors to the lack of representation are discussed in detail in sections 4.1, 4.2, and 4.3.

3.2.1.3 *Brand-specific considerations*

3.2.1.3.1 *Trial or demo versions*

When a potential user enquired about purchasing or acquiring a tool, the majority of responses recommended trying the demo or trial version. In some translators' opinions, downloading the trial version is "*the best option, this way, you'll be able to test it on actual projects*" (TC72-02). When asked for advice regarding tool choice, most authors, rather than promoting their preferred tool, chose to remind the potential user that in order "*to find out which CAT is more suitable for your work, you can download the demo version*" (TC02-04). These authors seemed to recognize the potential of the use of trial or demo versions to provide an indication of how a particular tool would suit a user's specific needs, a significant benefit which was highlighted in the literature review (section 1.3.3.1).

3.2.1.3.2 *Price*

The topic of the purchasing price of TM systems was one of the most commonly discussed issues within the corpus (a total of 87 postings addressed the issue). The examination of these postings revealed another divide among translators. Some believed that, while the cost of TEnTs is significant (describing it using terms such as "*huge amounts*" (TC13-05), "*small fortune*" (TC60-19), "*significant sum*" (P18-01), and "*hefty price*" (P42-24)), the tool was a necessary investment in quality (e.g. "*It's a pity that you have to invest so much money no matter what you want to undertake*" (TC64-01); "*I have to agree that Trados (and other CAT tools) tend to be expensive. But then again, they're professional tools, and, well, as translators we know that quality can cost money*" (TC29-02)). One author suggested that "*if you don't have any experience of CAT tools [...] try a less expensive tool, the*

financial risk being a bit lower then” (TC02-03). Others believed that TM systems are “*fairly priced*” (TC41-06), and some were relatively unconcerned with price, indicating that (at least some) TM systems are “*free*” (TC41-05). Twelve postings designated TM tools as worthwhile and/or necessary investments; however, nine authors stated that the use of TM tools was not beneficial to them.

A cross-examination of postings relating price to brand names allows for interesting observations. A total of 65 brand- and price-related comments were made, and can be interpreted in one of two ways: the tool is reasonably priced or the tool is expensive. MetaTaxis received 13 positive comments and no negative comments, suggesting that of the tools discussed in these postings, it was considered to be the most affordable and reasonably priced based on its quality. Conversely, SDL Trados was indicated as expensive in 30 separate postings, indicating that authors considered it to be the most pricey tool for the value discussed in these postings. Other opinions addressed Wordfast (10 for reasonably priced), Déjà Vu (two for reasonably priced, four for expensive), Star Transit (two for expensive), Across (two for reasonably priced), and MemoQ (two for reasonably priced). OmegaT was also discussed frequently, but, as it is free, no comments were made relating to whether the price tag reflected the value of the tool.

While some seemed sure of the investment, others were wary of the expense and worried about their return on investment (ROI), a factor that would most likely influence the amount a potential buyer would be willing to spend. Again, translators were divided on the issue of whether or not the ROI was guaranteed. Some believed that the tool will “*increase your productivity and earning power*” (TC01-04), whereas others were more hesitant, realizing that “*it could easily take a year or two before you see any real productivity gains. Add to that the time and effort in learning the tool, and you will see that the costs are not purely monetary*” (P18-03). While most potential buyers want a clear-cut answer for the question of their ROI, some authors directly related the potential for the tool to pay itself off to the type of translation work, the user’s desire to use it, and comfort with new software applications. Recommendations included: “*At your stage of business development your focus is better spent on doing your work as well as possible*” (TC01-02) and “*Buy it as and when you feel you are earning enough to justify it*” (TC01-02). Users who have had positive

experiences were also quick to share their opinions: “*The money I spent buying CAT tools came back to me in not more than 2 translations*” (TC60-06). In one situation, the author was considering purchasing a TM system for one particular job, and after providing some details about this job, a fellow member confidently replied, “*For a job like you mentioned I would get the license, it pays itself back right away*” (P08-02). Out of a total of 87 postings relating to price, five contained indications that the return on the investment occurred in the short-term, whereas six authors stated that benefits occurred within the long-term.

Some believed that if a certain tool is required by a client or agency, the client or agency should be paying the price. One author clearly stated his stance on the price of tools and the responsibility of paying:

But what happens with freelance workers, who are expected not only to run their business (and feed their family), but to also cover all costs related to tools and means of production? If a 2000 € licence fee per workstation can be a joke for a large company, it can be a serious burden for someone struggling to make a living on their profession of choice - especially since this costs is often incurred long before any potential jobs actually appear. Okay, yes, indeed, these are professional expenses, and, as such, are tax-deductible. But still, someone has to meet the bottom line. Now, methinks, shouldn't this cost be passed over to the intermediary, to these lovely agencies who are so deeply concerned with our welfare that they keep providing is with stimulating, top-quality jobs at great rates and with relaxed deadlines? As in, "Cool, so you want me to work for you. What's this thing you use? OK, no problem, give me and advance for a licence, then deduct it from me forthcoming invoices..." (TC29-14).

Whether a potential buyer considers purchasing a TM system to be a requirement or a well-thought-out and planned investment, price is without doubt a motivating factor, even perhaps the most significant.

As noted in the literature review (section 1.4.5), the benefits of using TM systems (e.g. increased productivity and translation quality) vary from user to user, and the amount of time needed to recoup the expense of purchasing a system largely depends on individual circumstances (i.e. familiarity with computers and TM systems, type of texts being translated). Section 3.2.2.5 contains further discussion of the benefits seen by ProZ.com and TranslatorsCafé.com members as related to procedural and textual effects of TM usage. (Also see section 3.2.1.1.1 for opinions related to repetition, text type, subject field, and client-specific projects and their effects on the usefulness of TM usage.)

The findings of three separate surveys indicated that perceptions of the initial cost and expected return on investment contribute to non-adoption: the results of the eCoLoRe survey showed that 51% indicated the initial cost and 34% indicated ROI as inhibitors to adoption (Wheatley 2003, n.p.); Dillon and Fraser noted that a high number of non-users expressed a “vague” interest in TM usage but indicated that the time and/or financial investment may be too high (2006, 73); and Lagoudaki’s survey revealed that 25% of non-adopters believed or knew that the initial cost was “too much” and 3% had previously used a TM system but found that there was no benefit to use (2006, 17). These findings and the contents of the corpus as discussed above indicate that both the initial cost of a TM system and the potential ROI are significant factors when choosing to use a TM system.

3.2.1.3.3 Updates or new releases

The discussions relating to updates or new releases pertained to two issues: new releases or versions of TM systems which must/may be installed, and new releases of OSs, or integrated word processors or text editors (i.e. other software directly affecting use of a TM system). As noted in the literature review (section 1.3.3.3), upgrades may be necessary in order to process and produce new file formats (as required by the client), to continue receiving technical support, or to ensure the TM system can function properly with the OS, or an integrated word processor or text editor.

Of a total of 49 update-related postings, 32 concerned new releases of OSs or word processors (specifically and uniquely Microsoft Office Word and Windows). This may be a result of the time period during which the postings were made, as Microsoft issued new versions of both Windows (Vista) and Office (2007) during this period. Almost immediately, many TM users found themselves struggling to make their software packages work properly. Although these new releases may have been anticipated, developers of TM systems still needed to develop service packs and updates, and in these instances, these authors found themselves unable to access their TM systems with their current OS or word processor version and wondering when a patch would be released. TM system developers found not only that the compatibility of systems was affected by these updates, but also that the new XML formats (e.g. .docx., .xlsx, .pptx) introduced

with Microsoft Office 2007 had to be addressed. In the meantime, some translators were temporarily unable to access these files or were searching for workarounds.

The issue of the price of updates to TM systems was addressed 19 times in the corpus, 11 of which revealed dissatisfaction with having to pay for new releases, and 8 of which supported free version releases. When the postings were analyzed based on brand, it was discovered that the majority of negative comments (nine of eleven) were made in relation to SDL Trados; the remaining two discussed Wordfast. The SDL Trados users were quite vocal, and comments included: *“I too began CAT with Trado\$. I used it steadily for a number of years, then found that there were too many bugs and too many requests for upgrades (and payments for said upgrades)”* (TC45-02); *“SDL have announced that any version including and prior to Trados 6.5 and SDLX 2004 will not be recognised as a qualifier for an upgrade to any current or new version after April 1st 2008. In spite of the excellence of TagEditor, I have about six major reservations about SDL as a company, but this announcement appears to mark a high point in its arrogance towards translators. Is it about time translators voted with their feet?”* (TC43-01); *“[There are] three possible options for current Trado\$' users concerning its new update policy: 1) get an upgrade [...] 2) tough it out with the old versions [...] or 3) say farewell to SDL and buy one of its competitors' products. I think that the 3rd option is the most probable”* (TC43-02); and *“[SDL has] a restrictive upgrading policy suddenly sprung on translators who bought in good faith”* (TC43-04). While the presence of these SDL Trados discussions may also have been a reflection of the time period of the postings (SDL Trados Studio 2009 was released in June 2009), one user acknowledged this issue more generally, noting that *“a lot of purely commercial software is designed so that you are locked into buying upgrades and so on, which seems maddeningly exploitative”* (P18-17).

Unlike that of the negative comments, the distribution of the positive comments was relatively even: three for MemoQ, two for Déjà Vu, two for Wordfast, and one for MetaTaxis. Regarding MemoQ, one user wrote, *“The licensing system of MemoQ is fine, as when you buy a licence, you get all future versions for free”* (TC41-06). Similarly, a contented Wordfast user noted that all updates and version releases were free within three years of purchasing a Wordfast licence (TC74-02), a reasonable time period for the user as the investment was worth it, and the fee after the three years was less than the purchasing price.

As noted in the literature review (section 1.3.3.3), the availability, timeline, and pricing policy for updates and new releases should be considerations for potential and current TM users. Although this issue may be under-represented in the literature, the analysis of comments in the corpus revealed that these issues are significant and that users have strong opinions related to vendors' pricing policies for new releases.

3.2.2 Using a TM system

3.2.2.1 Training

A total of 78 postings addressed training on TM systems and the learning curve involved. Of these postings, 21 contained either positive or negative references to the learnability and ease of use of seven different brands of TM systems and one in-house TM system (10 positive references, 13 negative references). The main contrast is evident in postings related to Wordfast; three authors made positive comments, and three made negative comments. The positive comments included: “*What I found imprerssive was the ease of learnig*” (P18-09), “[*Wordfast was*] *at once intuitive and easy to get started with*” (P18-12), and “[*Wordfast was*] *very easy to set up and start using*” (P18-17). Negative opinions were “*basically, it's just horrible to use*” (TC83-02), it was “*inconvenient and had a daunting learning curve*” (TC45-01), and that “*if you can learn how to use Wordfast, then Trados should not be difficult*” (P18-11). While these users of Wordfast were divided on the issues, Trados users participating in these 21 postings made almost exclusively negative comments (nine negative, one positive). The program was described as “*very complicated*” (P42-39), “*a big pain in the neck*” (P18-19), “*difficult to learn and very frustrating to use*” (TC60-05), and “*impossible to master without a guide [manual]*” (TC02-04). In fact, even the sole positive comment is not as enthusiastic as the others (one each for MemoQ, Similis, Déjà Vu, and MetaTaxis, and two for Across): “*I also believe that learning it isn't such a big pain*” (P18-20). While some tools may be inherently more or less easy to use, personal preferences and training are also important factors.

The following three sections further discuss issues related to training – time, availability, and price. These sections refer not only to guided training involving human interaction (i.e. workshops, training courses), but also individual and independent training

(i.e. through the use of manuals, tutorial, videos, webinars). (The issue of training is related to technical support and is discussed further in section 3.2.2.4.2.)

3.2.2.1.1 Time

Fifteen postings referenced the time investment needed for training. Of these, eight specifically commented that it does take a significant amount of time to learn how to use a TM system. Two authors recognized that time set aside for training could result in the potential loss of work. For example, one author wrote:

The 'hidden' expense that concerns me more is the time spent learning — first of all, the jobs lost if days have to be taken practising; or else, the loss of productivity on real jobs during the learning period. At the frenetic pace of my work, I can rarely afford the luxury of taking twice as long over a job just so I can use CAT, create a TM and / or a glossary. I have tried it on a few jobs, but the time taken slowed me down so much, in the end I went back to doing it the old way! Maybe I could learn to get faster — but I can't afford to lose all my clients in the meantime! (P18-14).

Another author stated that the time investment in learning how to use TM systems does not necessarily end when the training does: “*Although you can get the hang of CAT tools fairly quickly, it takes a lot of time and experience to get the best out of them*” (TC01-04). In addition, three authors indicated that this investment was not worth the perceived return and stated: “*It would take more time to learn than it would save*” (P18-01) and “*if you work on marketing documents or more creative texts [...] I would find that learning to use a CAT tool would be a waste of time*” (P18-04).

An examination of brand-specific discussions relating to the training time investment revealed five negative comments towards SDL Trados (e.g. “*If you buy Trados you will spend a long time learning it*” (TC01-02)) and one positive statement. Other brands discussed were Wordfast (one positive statement, one negative) and Across (one positive comment).

The training period, as indicated in the literature review (section 1.4.1.1), is potentially stressful and may result in decreased productivity (Bowker 2002, 115); many of the authors who discussed training recognized these consequences. Furthermore, the analysis of postings related to training revealed that some translators are choosing not to use TM systems, as the time investment for training was too great or the ROI would not compensate for time spent on training. These opinions reflect the findings of the

eCoLoRe survey, in which 38% and 34% of non-adopters cited the need for training and the difficulty associated with learning as inhibitors to adoption (Wheatley 2003, n.p.).

3.2.2.1.2 Availability

Methods of training discussed included individual study with user manuals (12 postings), tutorials (six postings), training videos (two postings), and information on vendor websites (one posting), as well as formal trainer-led courses or workshops (nine postings) and the SDL Trados certification program (29 postings).

Three authors commented on the overwhelming size of user manuals, specifically those provided by SDL Trados (e.g. “1000-page documentation” (P42-39), “300-page document” (P42-30), and “entire set of manuals, which amounts to around 700 pages” (P18-04). While five authors made positive statements about user manuals (e.g. the MetaTaxis manual was “*idiot proof*” (P19-03) and “*the best CAT tool manual*” (P42-07); the Wordfast manual was “*easy to follow*” (P18-07), “*well written*” (P08-03), and only required “*three days (max)*” to read (P18-09)), four authors did not find this method of training to be useful (e.g. “*could not make head nor tail of the manual*” (P18-07) and “*glossy manuals - obviously SO much care has been spent on them - which are fundamentally, monumentally useless for learning*” (TC43-05)). Four of the five negative comments were related to the SDL Trados manual and one was related to the Wordfast manual.

The postings pertaining to tutorials did not contain the strong opinions of those related to manuals; of the six postings, five simply referenced the availability of tutorials (Transit, Swordfish, an in-house TM system, and SDL Trados) and one contained a description of the Wordfast learning tools as “*completely incomprehensible*” (TC83-02). References to video materials were also neutral: one posting advertised an upcoming SDL Trados webinar (TC15-01), and the other posting contained a link to a Swordfish training video (P26-07). In the posting discussing website information related to Wordfast, the author stated, “*I would definitely like to learn more, but I am not quite sure how to go about it. I know there is a wealth of information on their homepage but at the moment it seems rather bewildering*” (TC70-01).

Nine postings mentioned training courses provided by the vendor: seven related to specific courses, and two generally discussed the availability of training courses. One

author noted that “*Trados will give you a headache at first, guaranteed. It is not user friendly. [...] But if you do it right (either buy training or read the entire set of manuals [...]) and stay patient, you will discover useful ways of using it*” (P18-04). Another author simply stated that “*Training courses are expensive, but strongly recommended*” (TC01-04). Both authors recognized the need for proper training in order to maximize the benefits of TM usage.

Twenty-nine postings discussed the SDL Trados certification program. The majority of postings (21) were within one specific ProZ.com thread relating to a job posting seeking a Trados-certified translator (P21). While the legitimacy of the requirement is another debate, four authors deemed the program to be worthwhile and demanding, although somewhat sarcastically in certain cases: “*isn't a Mickey Mouse test*” (P21-14), “*[it is] quite challenging*” (P21-21), “*it is good to have as many papers as possible*” (P21-17), and “*what better way to prove your skill than with a piece of paper*” (P21-23). Contrasting opinions were expressed by three authors, one who referred to the program as “*lunny*” (TC43-04) and another who described it as providing “*silly certifications*” (P21-04). One author in particular was especially displeased and wrote that SDL Trados sells “*worthless papers [...] at the price of gold as some kind of self-satisfaction proof*” (TC43-05).

The corpus contains no specific discussion related to training received in academic settings, aside from one author’s reflection: “*It seems that knowing how to use CAT tools is part of 'modern universitarian formation' if you want to become a translator [...] I think that, for future translators, being unaware of CAT tools can become suicidal*” (TC60-11).

As highlighted in the previous quotation, not only can a lack of training be detrimental for less-experienced translators, but it can also lead to the investment in a tool which cannot be used. As Lagoudaki’s findings indicated (see section 1.6.2.6), 16% of respondents had purchased a tool but did not know how to use it (2006, 17). As one author pointed out, owning a tool does not necessarily mean that the user knows how to use it (P21-16). While no authors explicitly indicated that they owned a tool but were *unable* to use it, one author did state, “*I've had Trados for 6 months and am yet to even open the program*” (TC01-09). The author gives no indication as to why the program remains unopened. This is open to speculation; however, a lack of training is a viable hypothesis.

In cases where a user has not received proper training, he or she may turn to discussion forums for help and guidance when troubleshooting. This form of support is discussed in section 3.2.2.4.2 as related to technical support.

It is interesting to note the lack of discussion related to training provided in translator training programs and by professional associations, as both of these sources were strongly favoured by TM users in the eCoLoRe findings: 68% of users believed that professional associations should provide training, and 33% “looked to a university for assistance” (Wheatley 2003, n.p.). This lack of discussion may be a result of the types of translators whose postings were included in the corpus, or a difference in preference of the provider of training; these possibilities are discussed in more detail in sections 4.1, 4.2, and 4.3.

3.2.2.1.3 Price

Only six postings related to the financial cost of training, five of which specifically addressed the high price of SDL Trados workshops in Poland and Lithuania (TC27). The only other comment relating to price was quoted in the section above, stating that while the cost of training courses is high, they are valuable (TC01-04). Sections 4.1, 4.2, and 4.3 discuss possible reasons for this lack of representation.

3.2.2.2 File formats

3.2.2.2.1 Input and output files

File formats of input and output files was one of the most discussed topics in the corpus (there was a total of 340 references related to source and target formats, files created by the TM system, and TM formats). (See section 1.4.2.1 in the literature review for explanations of these formats.) The majority of discussions related to the specific file formats a particular TM system could handle, create, and convert. These issues will not be discussed in detail as they tend to be technical and highly specific to certain tool brands, file types, and users.

File format support, compatibility, and conversion are addressed in the literature; however, the discussions do not focus on specific formats, but rather emphasize the need for potential users to determine their own specific needs and investigate the capabilities

of TM systems as related to their particular format requirements. This emphasis and lack of specific format discussions is warranted, as the context of TM usage and file formats used differ significantly from user to user.

3.2.2.2.1.1 Source and target file formats

Contributing authors were working with many file formats (e.g. desktop publishing (DTP) files, computer-aided design (CAD) drawings, OS and program files, web pages, XML files, spreadsheets, presentations, and word processor files) and were frequently researching how to manipulate their TM systems in their current states to read and write the necessary formats.

The most commonly discussed source and target formats were Microsoft Office formats (43 references), followed by XML files (33 references), DTP formats (produced by programs such as QuarkXPress or Adobe InDesign) (24 references), HTML files (19 references), system and program files (11 references), XLIFF files (11 references), .pdf formats (9 references), and CAD file formats (9 references).

Three of the nine surveys discussed in section 1.6 investigated the formats used by respondents. The eCoLoRe findings indicated that the Microsoft Office formats were the most commonly used (97% used .doc, 66% used .xls, and 45% used .ppt), followed by .rtf (78%), and HTML files (46%) (Wheatley 2003, n.p.). The results of Fulford and Granell-Zafra's survey also found word processor formats to be the most used for document production (by 99% of respondents) (2005, 9). DTP formats were also used by 17% of respondents (Ibid.). Finally, Lagoudaki's results indicated that 96% of respondents used Microsoft Office formats, 43% used .pdf files, 26% used XML or HTML files, and 12% used DTP files (2006, 12). While no conclusions regarding the amount of use of these formats by the contributing authors can be made, it is interesting to note the number of references to DTP and .pdf formats in the corpus as compared to the survey statistics: DTP formats were often discussed, whereas .pdf formats were seldom discussed. These differences may be due to the fact that authors may be more likely to seek help or discuss the formats that they seldom use or with which they are not comfortable working.

The postings relating to source and target text formatting (23 postings) within documents were also technical and specific to the message author's particular situation. The main issue was maintaining proper formatting (e.g. layout, text size, bold, italics) when translating and producing the target text within the TM system. Problems differed in type and severity based on the contents of the file, the TM system, and the text editor or word processor (either integrated or proprietary).

3.2.2.2.1.2 Legacy file formats

No postings in the corpus pertained to the formats of legacy files (see section 1.4.2.1.2 in the literature review). Sections 4.1, 4.2, and 4.3 provide a discussion of possible reasons for this lack of representation.

3.2.2.2.1.3 Files created by the TM system

A total of 96 postings contained references to the proprietary files created by TM systems during the translation process. These discussions are closely related to compatibility and are addressed in section 3.2.2.2.2.

3.2.2.2.1.4 TM formats

Seventy-six references were made to the specific file formats of TMs. These discussions are closely related to compatibility (section 3.2.2.2), open standards (section 3.2.2.2.4), vendor lock-in (section 3.2.2.2.3), and client format requirements (section 3.2.2.6.1) and are addressed in detail in these sections.

3.2.2.2.2 Compatibility among TM systems

The issue of compatibility among TM systems (i.e. compatibility not achieved through the use of open standards such as TMX) was a much-discussed topic in the corpus. All but six of the 46 postings related to compatibility between Trados and another brand of TM system (i.e. the user's preferred brand). This need for compatibility generally arose from client, agency, or outsourcer requirements (see section 3.2.2.6.1 for discussion); the translators were required to deliver their translation and other files in SDL Trados proprietary formats (especially the bilingual .ttx files (see section 1.4.2.1.3 of the literature review)). The postings discussed conversion between Trados and tools

such as MetaTaxis (eight occurrences), Wordfast (seven occurrences), OmegaT (six occurrences), Déjà Vu (four occurrences), Heartsome (three occurrences), MemoQ (two occurrences), and Across (two occurrences). Although some TM users were feeling pressured to use Trados, they were eager to find methods for delivering the necessary files while working with their preferred system. One user was confident in his approach and stated:

If you buy MetaTaxis, go ahead and tick the Trados box, because the agency would never know that you used MetaTaxis rather than Trados. <puts on innocent face> SURELY what the agency wants to know is whether you can read and write Trados TMs and work with that format, and not whether or not you have specifically shelled out the money for that program. (TC01-08)

The authors recognized that clients or agencies have specific file format requirements, but they did not necessarily agree that these requirements can only be met by using SDL Trados. Another user stated: “*That doesn't mean that you absolutely need TagEditor [an SDL Trados tool] to translate these TTX files. There are plenty of solutions around, and you can perfectly translate TTX files using Wordfast, DejaVu X, MemoQ, Metataxis or basically any CAT tool that would allow you to process tagged RTF files*” (TC21-02). And, according to one author, the use of a tool other than SDL Trados poses no issues: “*Not in a million years can an outsourcer determine, with what tool have you arrived at the final translation*” (P21-13). This, however, may not be the case. As one author stated, “*even if some people may state there are "compatible" tools, the truth is, that most of them still use Trados for pre/post processing their jobs. A 100% compatibility does not exist*” (P42-02). According to another author, the possibility of incompatibility makes it advisable to have an open relationship between the translator and client:

You should tell your client what you're using. Reassure him that you will be able to exchange TMs and send properly translated documents, but don't try to conceal the fact that you're not using directly Trados. Sometimes, quirks happens with "Trados compatibility", and when it does... you'd better have your ass covered by having been honest with the client. Sorting together the eventual problem is way better than delivering unusable files. One of the common culprits with TMs exchange when not 100% Trados compatible is that your client may very well get lower matches with a different tool (because of layout endocing difference, because of diffrent segment rules for TUs, etc.). I suggest you first try your alternative tool on a few files and see how things go and if the client can handle your deliverables. Better safe than sorry. (TC02-03)

While the literature does address the need for evaluation of individual file format needs and client requirements (Bowker 2002, 126), specific compatibility options and workarounds are seldom addressed. In García's 2003 investigation of Lantra-L messages, the author notes that systems were used as alternatives when clients required files to be delivered in a format produced by a system other than the user's preferred system. Specifically, García notes the use of Déjà Vu as a "Trados substitute": "Trados is, unquestionably, the choice of agencies; DéjàVu just seems to be that of freelancers. The assumption seems to be that when agencies ask freelancers to do a job 'with Trados', then DéjàVu (or Wordfast, or Transit, or SDLX...) will be adequate alternatives. Just how adequate, however, is a matter for debate" (García 2003, n.p.). Although specifics of Déjà Vu/SDL Trados compatibility were discussed in García's 2003 article, relatively little research has been conducted relating to precise compatibilities among systems and translators' strategies for achieving this compatibility. The issue of compatibility is closely related to brand and format requirements of clients and is discussed in further detail in section 3.2.2.6.1.

3.2.2.2.3 *Vendor lock-in*

The issue of vendor lock-in was addressed specifically 10 times in the corpus, all but two in relation to SDL Trados. (The other two comments discussed vendor lock-in in general.) Discussions were based on one of three issues: required payment for new releases (see section 3.2.1.3.3), a lack of support for open standards (in favour of proprietary formats), and perceived relationships between vendors and translation agencies.

In one discussion, originally addressing keyboard shortcut customization, one author commented "*And if they already bought the tool, the vendor's 'only' interest is to charge for updates, i.e. keep them locked in?*" (P12-03). Another author has noted the situation as a result of the user's inability to break free: "*Once you've bought SDL Trados, you cannot get rid of it. The licence stipulates that you cannot transfer your licence to anybody, which means you are not allowed to sell it. Which results in more licences gathering dust on the shelves of those who decided Trados was not good enough for their business - and more licences sold for SDL*" (P28-03).

Others blame a lack of incompatibility for the current situation, as one author wrote, “*My main reservation with SDL Trado\$ is that they try very hard to make their products incompatible, with proprietary formats and the like, and that they make buyers think that only Trado\$ products can work with TM's*” (TC43-05). Another author indicated that users are not the only uninformed parties in this situation:

Most any job that requires the production of translation memory files can be done using any CAT tool that is TMX-compliant. Therefore, Trados is not a must. [...] As for client asking for Trados - they are only asking for it because of the same misconception [...] and many think that the only CAT tool on the translator's end able to produce Trados compatible TMs is Trados. They are dead wrong - just like many a freelancer. The bottom line is you don't need Trados. (P28-13)

Finally, others indicated that the issue may be the result of clients’ brand requirements (section 1.4.6.1 in the literature review and section 3.2.2.6.1 in the corpus analysis) as seen in the following posting:

I get the impression that SDL offers something really valuable to translation outsourcers who post job ads now and then absolutely demanding the use of Trados, possibly in hoping that wannabes, newbies, and generally gullible people will buy it before attempting their first translation for cash. It seems to me the general message they want to convey is that it's utterly impossible to translate anything without buying Trados. (P21-25)

Another author echoed this sentiment by stating, “*My perception of this company [SDL] is that it hopes above all to sell to giant corporations, who will put pressure on translation agencies to buy, who will likewise pressure translators to buy*” (TC43-04).

Although the number of postings referencing vendor lock-in may not be high, it is interesting to note the recurrence of the issues surrounding SDL Trados in relation to the price for updates and new releases (section 3.2.1.3.3), compatibility among TM systems (section 3.2.2.2.2), and client brand requirements (section 3.2.2.6.1).

Vendor lock-in is not specifically addressed in the scholarly literature; however, the Déjà Vu compatibility guide does note that TM users should be cautious when choosing a tool in order to ensure that it can read formats created by other systems and access their resources in the future (ATRIL 1999, 2). The existence and consequences of vendor lock-in seem to be somewhat under-represented in the literature compared to the

number of postings in the corpus (and the strong opinions expressed therein) that discuss vendor lock-in and related issues.

3.2.2.2.4 *Open standard formats*

Although compatibility was among the most discussed topics, relatively few translators seemed to be turning to the open standard TMX to combat the issue, but rather were trying to find brand-specific compatibilities (i.e. using MetaTaxis, Wordfast, or Déjà Vu to process SDL Trados proprietary files). In fact, there seemed to be a significant number of authors (eight) who were unaware of or misinformed about the TMX standard. One user recognized the potential of TMX to aid in achieving compatibility, but was mistaken as to its origins and larger purpose: “*Purchase a [...] program that read and writes .tmx files (“Trados-compatible” as it is called)*” (TC02-02). Although some realized that if “*[you] make sure you’re using a tool that can export to TMX format [...] you’ll be OK and have a TM you can use almost anywhere*” (TC14-02), the misconceptions of TMX as being only a “Trados-compatible” format or an easy solution for “*programmers [who] don’t want to spend much time developing compatibility with lots of other tools*” (P16-08) indicate that awareness and comprehension of open standards and their purpose is somewhat low. In fact, SRX was only mentioned twice in the corpus. Those who indicated support for open standards and recognized their potential for facilitating compatibility recommended that “*translators, agencies and software developers should concentrate on TMX*” (P19-04).

This lack of awareness and/or understanding of TMX does not necessarily coincide with the findings of the LISA 2002 and 2004 surveys (14% unaware in 2002 and 20% unaware in 2004) (Lommel 2002, 15; Lommel 2004, 17) (see sections 1.6.2.1 and 1.6.2.2), and may indicate that awareness has not improved in the past few years. There is a difference between ‘awareness’ of TMX and comprehension of its purpose and use; the LISA surveys investigated ‘awareness’ of TMX, but ‘awareness’ does not necessarily indicate that a respondent understands the principles and usage effects of open standards. Furthermore, it should be noted the types of users whose postings were included in the corpus may differ from those who responded to the LISA surveys. As noted in section 1.6.3.2, the respondents to the LISA surveys are likely to be involved in the localization

industry, and as such, the results of the LISA surveys may reflect the highly technical nature of TM use in the localization industry. This may have had a direct influence on the level of TMX awareness indicated in the findings. See sections 4.1, 4.2, and 4.3 for a discussion on the relationship between the types of users and the topics addressed in the corpus.

3.2.2.3 Matches proposed by a TM system

3.2.2.3.1 Types of matches

The corpus contains 12 references strictly related to the types of matches proposed by TM systems. Half of the postings pertained to technical problems encountered when retrieving matches, especially with tagged file formats. One thread (TC22) was related to the types of matches indicated in OmegaT's analysis results. In the remaining posting, the author stated that “*Trados' search functions are unbelievably poor (for example, a translation unit has to be at least a 30% match to show up anywhere [...])*” (P42-21).

While this may seem to be insignificant representation based on the issues described in the literature review (section 1.4.3.1), discussion of matches is most often tied to other issues of TM usage (usefulness of matches (section 3.2.2.3.2), payment (section 3.2.2.6.2)), and the buying/selling/trading of TMs (i.e. using another translator's TM) (section 3.2.2.6.4)), and it is in these contexts that the majority of match-related discussions occurred.

3.2.2.3.2 Usefulness of matches

The eight postings pertaining to the usefulness of matches related to one of two points: the scope of the TM (e.g. one large TM versus smaller, more specific TMs) or the source of the TM (i.e. personally created or received from an agency, client, or colleague). One author asked whether “*it is advisable to set [create] several TMs in a CAT system ([...] in relation to your specialization) or whether it is more logical to put everything in the same basket making it therefore only one TM*” (P39-01). On this issue, two authors indicated that they preferred to use only one TM (an “*ultra-large*” TM (in order to achieve high daily output) (P31-01) and a “master TM” containing segments from “*agency clients willing to share*” (in order to maximize the time benefits of TM systems) (P42-41)). One author

stated that he preferred to create smaller, more specific TMs (“*medium-sized*” TMs as “*I find that ultra-large TMs give too many matches, so it is hard to find good ones*” (P31-02)). Another author expanded on his preference for numerous smaller TMs, and wrote, “*As a freelancer, [...] I would rather create customer-specific TM's than specialism-specific TM's*” (TC39-02).

The second issue, that of the source of the TM, was mentioned twice in the corpus. Both authors indicated that the use of another’s TM is not always beneficial. One author stated that he chooses not “*to deal with the intellectual property of third parties (i.e. more or less reliable TM's provided by agencies)*” (TC60-02), and the other author wrote, “*I know when I get other people’s memories they can be of limited help*” (TC60-19).

As noted in the literature review (section 1.4.3.2) and demonstrated by the corpus references, this preference is personal; some users prefer to view all possible matches by using a large TM, while others prefer to limit their TMs to specific clients or subjects. This evidence supports the assertion in the literature that dealing with another’s TM is not always beneficial (Bowker 2002, 123). For further discussion of this issue and its impact on the quality of matches, see section 1.4.6.3 related to ownership and section 1.4.6.4 related to buying/selling/trading TMs.

Despite the importance given to fuzzy match values in the literature (O’Brien (1998), Bowker (2002), Fifer (2007)), the corpus contained no discussions pertaining to users’ fuzzy match value preferences. As with the division of TMs, this preference is most likely affected by the user’s particular context (i.e. the contents and source of a TM, the source text, or preferences for editing versus translating), and the lack of discussion may be attributed to the personal and individual nature of this issue. Additional possibilities for the absence of this issue in the corpus postings are addressed in sections 4.1, 4.2, and 4.3.

3.2.2.4 Technical issues

3.2.2.4.1 Customizability of TM systems

As noted in section 1.4.4.1, discussions relating to customization will focus on the display within the TM system and keyboard shortcuts (and not programming alterations

to the source code). No postings in the corpus pertained to the customizability of the display. Sections 4.1, 4.2, and 4.3 discuss possible reasons for this lack of representation.

As noted in the literature review (section 1.4.4.1), the use of keyboard shortcuts (as opposed to using menus, toolbars, and dialogue boxes) is a personal choice; some users prefer to execute the majority of commands using shortcuts (e.g. Ctrl+S), while others prefer using the mouse. Users who prefer shortcuts may appreciate the option of creating new or changing existing ones. One discussion relating to the customizability of shortcuts led to the involvement of a developer of MemoQ who explained why this customization may or may not be available in a given tool:

The lack of shortcut customization is either [...] a sign that the tool did not reach a level where this is the next logical step to take (e.g. it's more important to implement other functions), [...] or] competition, trying to prevent users from using other tools. The tricky thing, however, is this: if users can't use their previous shortcuts, they lose productivity, if they lose productivity, they won't change to your tool without a very special reason. (P12-03)

Based on the development and competition issues outlined in the above posting, those who use keyboard shortcuts frequently and desire customizability options may have to investigate this issue in-depth to find a tool that meets their needs.

Customizability options of TM systems are not addressed in the scholarly literature, which may be due to the brand-specific nature of available options. However, it should be noted that seven references to customizability were observed in the analysis of vendor promotional materials (see section 1.5), perhaps indicating that vendors are aware of the appeal of customizability options in TM systems.

3.2.2.4.2 Technical support

A total of 39 postings discussed technical support for TM tools. The most common form of support discussed was e-mail, but other sources of vendor-provided support such as telephone assistance, help documents, websites, and knowledgebases were also mentioned. Non-vendor-provided sources of support discussed were mailing lists, user groups (such as Yahoo! Groups), and other forums. Users indicated no preference for the format of technical support; however, in most cases, users attempted to find the information themselves before formally contacting customer service.

An analysis of postings relating to technical support revealed contrasting opinions based on the brand of the system. MetaTaxis support received three positive comments, being noted as “*unsurpassed*” (TC01-06) and “*first-class*” (P42-29), with two mentions of personal involvement of the developer. One author wrote, “*You can always count on Hermann’s help, if needed*” (P42-11). Across was discussed positively three times, with reference to “*very quick*” e-mail responses (TC13-05) and staff who are “*very polite*” (TC13-10) and “*nice*” (TC13-11). Three positive statements in relation to MemoQ’s technical support were made; support was described as “*excellent*” (P18-22), “*outstanding*” (TC45-14), and “*blazing fast*” (TC45-14). One author wrote, “*The developers of MQ were very responsive, helpful and really did their best to solve the problems very quickly*” (TC45-01). In fact, as further evidence of this fact, one of the MemoQ architects participated in the forum with the aim of aiding MemoQ users, and stated that “*We appreciate all your feedback, and we definitely keep an eye on threads like this*” (TC45-09). Similis and Wordfast received one positive comment each: “*The Similis people are actually HELPFUL; now that’s a refreshing change*” (TC05-01) and “*[At Wordfast] they’re pretty human, for a software company*” (P30-04).

The support provided by SDL for SDL Trados, however, was not viewed positively, receiving eight comments, all of which were negative. SDL support was described as “*subpar*” (TC29-18), “*notoriously slow*” (TC31-01), and as leaving “*something to be desired*” (TC37-03). Furthermore, users were not enthused about having to pay for this support, perhaps believing that it should be part of the cost of the tool. One SDL Trados user wrote, “*SDL Trados seems to think that customer service is important only if you’re willing to pay \$100 a year for the honor. Otherwise, they won’t even let you contact them*” (P42-19). One author stated, “*Generally, the more you pay for a product, the more support and development there is behind it (eg Trados, Deja Vu)*” (P18-17). However, considering SDL Trados was believed to be expensive by the authors of 30 postings (see section 3.2.1.3.2), this relation does not prove true based on the information in the corpus. (Support for Déjà Vu was not discussed.)

Although there was no specific discussion of support response times which compared open-source and closed-source tools, one author noted that OmegaT’s “*lack of guaranteed support*” was a potential issue as the project is managed by volunteers (TC83-

07). However, as no negative or positive comments related to OmegaT, this statement can be neither proved nor disproved within the context of the corpus.

In cases where TM users cannot or do not receive the necessary support (or may not have received sufficient training), they may turn to the advice and help of their peers on discussion forums, or in mailing lists or distribution groups. A total of 156 postings related to technical problems, yet less than one-third discussed vendor-provided technical support. This may indicate that discussion forums are seen as an alternative source for resolving technical issues and troubleshooting. While this may be one of the purposes of discussion forums, contributors became frustrated when users, who could not be bothered to learn the basics of the tool or troubleshoot a problem individually, asked simple questions requiring simple answers. In these instances, respondents may suggest that the user ‘read the friendly manual’ (RTFM). Two users in particular asked for help for what, in the first case, was described as “*questions about very basic usage*” (TC34-02), and in the second case required “*a simple answer*” (TC57-02). In the first situation, the original author (i.e. the one with the problems) was asked if he or she had “*read the documentation of an almost 1200 € package [...] you bought, or do you just assume it works on its own and it is so intuitive that you'll get to master it in only a couple of hours without actually trying to start any learning process whatsoever?*” (TC34-02). In the latter situation, the simple answer was described as “*a typical RTM case*” (TC57-02). As noted with regard to training in section 3.2.2.1, members of these forums demonstrated a willingness to help, but appreciated when a user also showed initiative to investigate the issue on his or her own.

An investigation of response time by fellow discussion board members revealed that the majority of authors seeking help on the discussion boards received a response within one day; in fact, 20 authors received a response within the same day of their posting. Of the authors seeking advice for technical problems, seven who posted on TranslatorsCafé.com did not receive any response, while only one author on ProZ.com did not receive a response.⁶²

⁶² A total of 16 authors received no response on TranslatorsCafé.com, whereas only two authors on ProZ.com received no response. Although no conclusions regarding the probability of receiving a response on ProZ.com versus on TranslatorsCafé.com can be made, it is interesting to note that in the specific threads included in the corpus, ProZ.com members responded to postings more often than TranslatorsCafé.com members. This, however, may be a result of the methodology (i.e. random sampling) used for thread inclusion or may be directly related to the content or topic of the authors’ postings.

The scholarly literature does not address technical support in any detail, aside from general discussions in relation to the time investment needed to use a TM system. Lagoudaki's survey included a section in which users rated aspects of their primary TM system brand on a scale of one to five (one being "not satisfied at all" and five being "excellent") (2006, 28). In the customer support category, Trados received a rating of 2.5, SDL received a rating of 2.6, and both Déjà Vu and Wordfast received a rating of 3.7 (Ibid.). The opinions in the corpus postings related to SDL Trados support were not positive (eight negative comments were made), indicating a somewhat negative image of SDL Trados support; Lagoudaki's findings also revealed a mediocre reputation for SDL Trados support, as its rating fell between "poor" and "quite satisfied" (Ibid.). No comparisons between the corpus contents and Lagoudaki's findings can be made for Déjà Vu or Wordfast, as support for Déjà Vu was not discussed and only one comment was made in relation to Wordfast technical support.

3.2.2.5 Procedural and textual effects of TM usage

3.2.2.5.1 Speed

A total of 21 postings discussed the effects of TM usage on the speed of translation. The discussions related, for the most part, to repetition, training, and ROI and have already been addressed in those contexts. (See section 3.2.1.1.1 relating to text characteristics, section 3.2.2.1 on training, and section 3.2.1.3.2 related to price and return on investment.) It is important to note that authors recognized the relationship between speed of translation and their specific situational circumstances (i.e. text type, repetition, familiarity with TM systems).

3.2.2.5.2 Quality

Eleven postings directly referred to the quality of translations produced using TM systems. On one hand, four authors indicated that the use of TM systems allowed them to ensure and improve the consistency of their translations. On the other hand, two authors contested this association between TM usage and quality. One author noted, *"I have always considered any CAT tool nothing more than that - a tool. It has nothing to do with style, with translation quality, with anything like that. A bad translator using a CAT tool is still a bad*

translator' (TC48-08). Another author suggested that use of a CAT tool may in fact allow unqualified translators to continue receiving work: *"I have often been asked to go back and manually edit/correct documents previously 'translated' using CAT — and the poor quality of those 'translations' might lead one to wonder if certain translators (including on this very site!) don't hide poor translation quality behind CAT productivity"* (P18-12).

As discussed in the literature review (section 1.4.5.2), information in TMs must be accurate and of high quality in order to produce accurate and high-quality translations (Austermühl 2001, 140). If not, a user may need to spend a significant amount of time reviewing matches to ensure they are translated correctly (see discussion relating to the usefulness of matches in section 3.2.2.3.2).

3.2.2.5.3 *Modifications to the translation process*

Six authors indicated that the use of TM systems had a positive effect on their translation process. Four of these authors preferred translating within the TM system. One author noted:

There is another advantage in CAT tools, rarely mentioned but to me very persuasive. That is simply the display. Trados for example goes through a text sentence by sentence, displaying the source sentences one by one in a coloured box with another box immediately below for your target translations. I find this much better than traditional methods of translating (over typing the source text, or writing the target text in a new document). (TC01-04).

The other authors indicated that TM use was beneficial to maintain the format (TC48-01) and to reduce the amount of toggling between windows while translating (P42-21). These benefits exist regardless of text type, repetition, and client specificities, and it is these types of users who use TM systems to translate all texts, rather than just repetitive texts.

Conversely, one author indicated that TM systems added additional steps to the translation process:

Instead of 1) seeing a source sentence, 2) taking the time to think about the words and phrasing I want, and then 3) producing a target segment, I now have to:
a) review the source segment
b) examine the suggested/matched segment

- c) examine each of the alternative segments (suggested and lower matches), especially to consider whether the translation is correct (other translators often pop up in the TM) and/or the particular phrasing fits*
- d) consider whether the matches contain the appropriate terminology for the context/client*
- e) even in the case of a 100% repetition, I still have to examine the match for context/terminology (and even more if I was not the original author)*
- f) potentially "fix" a fuzzy match suggestion to produce the target segment - and as people have pointed out, "proofreading" can often take more time than translation-from-scratch! (P42-41)*

The literature review contains no discussion of these types of procedural benefits; however, as noted in section 1.7, usage of TM systems for the translation of non-repetitive texts has been observed (García 2006) as a result of developing EBMT and sub-segment matching technologies. The literature does discuss procedural drawbacks in relation to TM usage (e.g. the need for and time involved with database management (Austermühl 2001, 140) and alignment verification (Bowker 2002, 11)); however, no postings in the corpus directly addressed these issues (although many referenced the time investment of using a TM system).

3.2.2.6 Client relations

3.2.2.6.1 Brand or format requirements

The brand or format requirements of clients was one of the most discussed topics in the corpus, with references to the issue in 138 postings. As discussed in section 3.2.2.2.2, many TM users believe that the choice of which tool to use should be personal and not the result of client brand requirements. A total of 103 postings related to client requests for SDL Trados use or formats (six comments relating to client requirement of Fortis, four regarding SDLX, and one each requiring Sisulizer, Lingobit, and Idiom Worldserver). Some non-Trados users feared that the number of jobs available to them and their chances of obtaining those jobs would diminish based on their choice. For example, one author posted, *"I am now worried that, despite my experience, I will just be disregarded by other potential agencies/ outsourcers for not working with TRADOS"* (TC01-01). In fact, one author believed the situation to be so extreme that *"most companies will not take you if you don't use TRADOS"* (TC08-01). Another author stated that *"the client has a right to*

demand specific RESULTS. If these results can only be obtained with a specific tool, then so be it' (P42-31), but that the client has no right to demand the use of a specific tool.

Furthermore, authors who were required to invest singlehandedly in the tool because it was a requirement of the client expressed their displeasure. Comments such as *"So, why is it the translator who is required to learn to use buggy software, pay for it, and put in extra work? How hard would it be for an agency to align the source and target documents and create a TM themselves when they work with a translator who doesn't use their CAT tool?"* (P42-21) and *"If an agency wants you to invest about \$1000 out of YOUR pocket for their own benefit, then there's something wrong with this picture"* (P42-04) indicate that some TM users believe that the translator should not be solely responsible for purchasing the tool. Others suggested that agencies should help the translator in making the financial investment. One author described the model used by Similis: *"What happens is that agencies that use Similis can send translations in 'kits' to translators; to use this kit there's a free download so you don't even have to buy Similis at all"* (P04-02). Another author negotiated tool price and payment with the agency:

They wrote to me saying that they would like to work more intensively with me, but for that I would need to own Trados, as their clients frequently insisted on its use. They had negotiated a 15% discount on the Freelance package, and advanced me the money against future invoices. The increase in volume this generated meant that I was free of the debt after just two weeks. (TC29-15).

Currently there is no universally satisfactory solution, and workarounds are being used or translators are simply accepting only the jobs for which they have the required tool and can supply the necessary formats. Although SDL Trados appears to be the industry standard (based on the contents of the corpus – see also sections 3.2.2.2.2 and 3.2.2.2.3 relating to compatibility and vendor lock-in and section 1.6.3.2 comparing the results of brand-name focused survey questions), a large number of translators are not necessarily content with this situation.

The issue of client brand or format requirements is identified in the literature, and the impact of these requirements is addressed in the findings of the eCoLoRe and Lagoudaki surveys: 41% of TM users who responded to the eCoLoRe survey were required to use TM systems by their clients (Wheatley 2003, n.p.), and 29% of TM users who responded to Lagoudaki's survey were required by an agency or client to use a TM

system (2006, 19). While the majority of TM users have personally made the choice to use TM systems, these findings and the discussions in the corpus indicate that a significant number of translators are required to use TM systems by certain clients or agencies, and these requirements play a significant role in adoption.

3.2.2.6.2 *Payment*

Of the 42 postings relating to payment, 30 contained opinions indicating that discounts should not be given. These authors believe, as one wrote, that a translator should “*apply your normal per word rate considering the content specialization/difficulty and requested timeframe. And do not forget to inform your client that you will not ask them to share the huge amount you paid for Trados to improve the translation quality, especially the translation consistency, i.e. the cost you PAY for your client's VALUE OF MONEY*” (TC16-02). Another author expressed her dissatisfaction with the idea of being paid less for matches in a TM: “*I'm pretty bitter about being *constantly* asked for discounts based on TM repetitions and fuzzy matches. As I tell my clients now - and am beginning to lose work because of this - repetitions and fuzzy matches do not necessarily lead to a reduction in the amount of time I spend on a translation, and in many cases actually add more time*” (P42-41). Conversely, some users disagreed with the belief that the use of matches reduces the amount of time needed to translate. One author noted, “*I have found that checking up on matches and repetitions often cause the same amount of work, or more, than translating from scratch does; after all, I am responsible for the entire translation, including the contextual appropriateness of translation memory matches*” (TC69-02).

For the most part, users seemed to be perplexed by the promotion of TM systems as tools that will increase income and productivity yet come with the expectation of lowered rates for matches. For example, one author wondered, “*What I do understand is that I am paying for the program. I am spending the time to learn the program, and the agency or the end-client wants heavy discounts plus a few free goodies; discounts and goodies not available without the expensive CAT that I paid for. Could be early Alzheimers ... but I don't see where my benefits come in...*” (TC60-05). Another author wrote, “*There is obviously downward pressure on rates due to CAT tools. It's ironic that we're expected to shell out \$1,000 or more, then offer discounts to boof*” (TC60-18).

Opinions differed as to how this situation came to be, some placing the blame on agencies or vendors:

I would place the blame on the aggressive marketing policy of the leading supplier of TM-based CAT programs. Their approach is ingenious: convince the end-user that insisting on their software could result in uniformity and lower prices, and then trying to push freelancers into using the software, promising an increase in turnover [revenue], while incorporating a feature which is designed to do exactly the opposite (the analysis of matches, partial matches and so on). (TC60-03)

Another author saw the situation as the result of an unclear path for the industry when establishing standards for payment (e.g. by the word or by the hour): “*We are standing at an exciting point in the industry — where there is a clear need for a new direction but no clear direction in which to head. In other words, the industry is searching for leaders, which is why the technologists have had so much sway lately*” (TC60-27). One author offered a possible solution and suggested, “*I refuse to offer Trados rebates - and so should you. It is the only solution we have, but if used consistently, it will work*” (P28-03).

These opinions are consistent with the issues outlined in the literature review (see section 1.4.6.2): translators are generally dissatisfied with the payment rates for TM matches, and payment practices may be a factor contributing to non-adoption (Wheatley 2003, n.p.). The debate surrounding the amount of time needed to review matches is still open (although it is clear that some time is required), especially as there is relatively little empirical data related to editing time and TM matches.⁶³

3.2.2.6.3 Ownership

In addition to tool and discount demands, some clients also require that translators turn over TMs or uncleaned bilingual files. A total of 12 postings in the corpus addressed this issue. Four separate authors indicated that they had no problem submitting the TM to the client. One author noted, “*Obviously I made it a separate TM to respond to the Agency*” (TC39-03). Another author had no reservations about providing her TMs, but recognized that TMs supplied by others may be of limited help (TC60-19) (see section 3.2.2.3.2 on usefulness of matches). A third author had no issue, but stated, “*They [the client] would have to recognize that the benefit of that goes to them, not to me*” (P42-34).

⁶³ One exception is a recent study by O’Brien which examined the processing speed and cognitive load of different match types through eye-tracking technology (O’Brien 2006).

Two authors indicated that providing the client with the TM may be somewhat problematic. One author stated that he was willing to fulfill such requests, but this would require additional payment: *“Handing over TM's [...]? That comes with additional charges [...] they are my intellectual property for which they don't pay... and which they will hand over to the next translator dealing with the next project for the same client!”* (TC60-06).

The literature review indicates that this may be a highly contentious issue, yet the corpus contained very little discussion pertaining to ownership and no discussion of copyright implications.

3.2.2.6.4 Buying/selling/trading TMs

The corpus contains little mention of the possibilities for exchanging TMs (only six postings) outside of providing and receiving TMs for specific projects. All six postings in the corpus related to the formation of the TAUS Data Association (see section 1.4.6.4). Four of the authors were not pleased with this development and expressed opinions such as *“I take it as a marketing gag. They know their TMs are not worth anything, but such a site may attract customers, like the free goodies sometimes distributed by retail shops”* (TC60-21), and *“When naive end clients approach me with ridiculously low rates expectations, I use to recommend them a FREE translation by Google. When the TAUB [TAUS] [...] project will be established, I can be even more helpful by recommending them to search whether their text is already translated”* (TC60-24).

Authors in the corpus were not discussing exchangeability options for themselves, but discussing how clients and agencies would benefit from buying, selling, or trading TMs. This lack of enthusiasm parallels the findings of the LISA 2002 and 2004 surveys (Lommel 2002, 16; Lommel 2004, 16), in which respondents showed little interest in the idea of a TM marketplace.

3.2.2.7 Non-adoption attitudes

The number of postings by non-adopters of TM systems is surprising: in fact, 44 postings discussed non-adoption of TM systems. These authors seemed to be informed of the current situation with regard to TM usage, and often reassured potential users that TM adoption is not mandatory or even necessary. Many validated their non-adoption

based on the type of translation work they perform, giving justifications such as “*most of my work is the kind of documents with zero repetitions where a CAT tool is more likely to slow me down*” (P18-12). Many struggled with their decision as they felt pressured to purchase and use the tools (see section 3.2.2.6.1 related to client brand and format requirements). One author confessed, “*I actually prefer working without CAT tools, but people seem to think you are mad if you say that*” (TC01-03). Other comments aimed to negate the idea that the use of CAT tools make a translator a professional. One translator appeared exasperated by statements linking professionalism with CAT tool usage: “*Having a CAT tool, or not having one, does not make you a more or less sensible person. It's just a matter of whether you need it or not for your business goals. It's sensible to have the tool if you need it. It is not sensible to pay for something you don't need. Period*” (P42-42). Regardless of the motivations behind non-adoption, the involvement of non-users in the discussions provides a somewhat balanced view for potential users.

The literature presents numerous factors contributing to non-adoption, and these have been previously discussed in comparison to the corpus contents in the sections relating to text characteristics and format (section 3.2.1.1.1), operating systems (section 3.2.1.2.1.2), price (section 3.2.1.3.2), training time (section 3.2.2.1.1), and payment (section 3.2.2.6.2).

3.2.3 Comparison with vendor promotional materials

Based on the analysis of vendors’ promotional materials described in section 1.5, certain comparisons and contrasts can be made with the contents of the corpus. While the issues addressed are similar, the frequency of references (and therefore the assumed importance) in the promotional materials versus in the corpus does not always correspond, indicating that TM users and vendors may not necessarily consider the same issues to be equally significant. Recommendations for bridging this gap are made in section 4.2.

In both the corpus and the promotional materials analyzed, the issue of file formats was the most commonly addressed. Although no specific details of file formats were tracked during the analysis of vendor promotional materials, the attention paid to file formats by vendors indicates that they are aware of the importance of this issue for

TM usage and are focusing on file format capabilities as a selling point, an appropriate focus as source and target file formats were discussed in 23.4% of the corpus postings. This awareness is further reinforced by the placement of compatibility as the fourth most commonly addressed issue; vendors recognize that users are not using one system exclusively and that the appeal of a given TM system is increased by its compatibility with other TM systems.

Further comparison between the volume of discussions and the references in promotional materials revealed that vendors were promoting features of TM systems that were not heavily discussed in the corpus. The analysis of promotional material showed that vendors tended to emphasize the technical aspects of their TM systems (i.e. features and functioning) rather than the benefits of using a TM system (i.e. increased quality and speed of translation, increased income). While these issues were discussed in the corpus, the volume of postings relating to ROI and the decision to purchase a tool was significantly higher than the level of emphasis placed on these issues in the promotional material analyzed. These discrepancies primarily affect new or potential users, as they are the authors enquiring about ROI and benefits of TM usage. Current and established users may already have experience-based opinions on ROI and benefits, and therefore may be more likely to investigate file formats and compatibility. The significance of these issues in the promotional materials may also indicate that vendors are aware of the role of client brand and format requirements in choosing a TM system.

The comparison of the analyses with regards to payment and increased income also allows for the observation of an interesting paradox. The promotional materials analyzed indicated increased income as a benefit of using TM systems; however, nowhere in the vendors' materials are there references to payment practices in the industry, an expected absence given the purpose of advertisements and websites. As noted in section 3.2.2.6.2, translators may find this situation to be ironic: vendors claim that the use of a TM system will increase a translator's income, yet translators find that they are sometimes required to lower their rates for matches found in a TM. One thread specifically addressed the issue of vendors' awareness of payment practices. The author who began the thread shared the subject line of an SDL Trados promotional e-mail: "*Translate more for less*" (P28-01). The phrase 'more for less' was interpreted by the

author (and subsequent authors who participated in the discussion) to mean ‘translate more *texts* for less *money*’, and these authors were discussing the irony of the situation. In a later post, one author suggested that this e-mail was not intended for translators, but rather for language service providers and agencies, thereby contextualizing the “more for less”: by requiring the use of SDL Trados, agencies and LSPs can increase their translation volume while decreasing the cost of translation by charging less for matches found in TMs (P28-05). Discussions such as this one may indicate that vendors are aware of current payment practices and TM users’ opinions and subsequently adapt their promotional materials to better address the target audience (i.e. a translator/TM user versus an agency, client or LSP). While no conclusions can be made in regards to whose ‘side’ of the issue vendors may be on, it is interesting to note that the benefit of “Lower Translation Costs” was referenced more frequently than “Increased Income” in the promotional materials analyzed.

Depending on the type of user and context of adoption, a translator may find that the information in vendor promotional materials does not adequately address the issues that are important to him or her, and he or she may have to investigate other complementary sources (e.g. discussion boards, mailing lists, scholarly literature) to obtain a more complete overview of a particular brand.

3.3 Concluding remarks

In this chapter I have summarized and highlighted the viewpoints and attitudes found in the corpus. While TM users were in agreement on some issues, the analysis of postings relating to other topics revealed varying opinions. In the conclusion I will provide my interpretations of the corpus data, discuss possible areas for further research and investigation, and evaluate the methodology of the project.

Conclusion

As stated in the *Objectives* in the introduction (section 0.2), the purpose of this thesis was to investigate translators' unprompted opinions of TM systems and the issues surrounding their use through the identification of a primary source, and subsequent collection and analysis of the data. These steps were accomplished through the implementation of the methodology described in Chapter 2. After completing the data collection and classification process, synthesizing the corpus contents, and comparing the contents of the postings to the literature review and vendor promotional materials, I can now make observations with regard to the contributing authors, interpret the contents of the corpus, comment on the lessons learned from this investigation of translators' attitudes, opinions, and perceptions, offer suggestions for further research and recommendations to assist the various stakeholders in achieving a better understanding of TM-related issues, and evaluate the methodology. (Interpretations and recommendations for further research are made using the thematic structure of Chapters 1 and 3; see Appendix I for a summary of my recommendations organized by stakeholder group. Appendix A contains a listing of the issues addressed in Chapters 1 and 3 and the conclusion and provides cross-references to these discussions within the thesis.)

4.1 Observations about contributing authors

Before interpreting the representation and content of the corpus discussions, it is necessary to make a few observations in regards to the type of translators/TM users whose postings were included in the corpus. The aim of this project was to investigate the opinions and attitudes of *translators*, without limiting the sample based on experience (as a translator and/or TM user) or employment status (e.g. freelance translator, in-house translator, member of management); therefore, the type of user (or non-user) was not taken into consideration during the data collection stage. Furthermore, it was not possible to identify the employment status and level of experience of ProZ.com and TranslatorsCafé.com members, as this information was not readily available and such an investigation would involve a significant investment of time.

As a result, the corpus contains postings from a wide variety of translators, including both TM users and non-users. However, based on the content of the postings and the amount of discussion (or lack thereof) related to certain issues, a few observations can be made. Firstly, it seems as though the majority of authors were TM users, although 44 postings contained the viewpoints and opinions of translators who were not using TM systems. Many of these non-users seemed to be informed about TM systems and the issues surrounding their use and, for the most part, had previously used TM systems in their work and subsequently discontinued this use. Thus, it can be assumed that the majority of authors were potential, current, or past TM users and were currently considering TM usage or, at one point in time, had considered TM usage.

Secondly, it can be assumed that the potential users who were investigating TM systems had already completed a preliminary evaluation of the benefits and drawbacks to TM usage; therefore, it is likely that those who had automatically dismissed the use of TM systems were not participating in the discussions.

Thirdly, based on the contents of the discussions and the issues being discussed, I believe that the majority of the contributing authors were freelance translators, including both those working directly with an agency and those working independently of an agency (i.e. with directly clients), and not in-house translators, clients, or agency employees. Specific examples upon which this observation is based are discussed in relation to the particular issues in the following section. Moreover, differences in work environments of freelance translators and in-house translators may affect discussion board use; in-house translators often work as part of a team and are most likely supported by a network of colleagues, whereas freelance translators often work alone and may not find opportunities for communication with other translators as readily as in-house translators. As a result, I believe that freelance translators are more likely to seek advice and guidance from their peers on discussion boards.

Finally, it should be noted that TM users at different stages of use may have very different attitudes, opinions, priorities, and levels of awareness; for example, an issue that may be of significance to a potential user may already have been addressed by an experienced user, or an issue may be of more importance to experienced translators than to recent graduates from a translator training program or individuals who have just begun

their careers in translation. Again, specific correspondences between the stage of TM usage and importance of an issue are addressed in the following section.

4.2 Interpretation of corpus contents

4.2.1 Purchasing or acquiring a TM system

Over 30% of the postings in the corpus related to purchasing or acquiring a TM system, indicating that many potential users were turning to discussion forums as a source of help and advice. Potential users who are turning to discussion forums should evaluate the advice given by both users and non-users in order to obtain a balanced view. While the majority of the participants in TM-related discussions were users, it was reassuring to note the number of non-users who also participated in the discussions. Continued input from both users and non-users will aid potential users in making their decisions based on information from both groups, rather than relying only on positive (or negative) feedback from current users.

Potential users often seemed hesitant to invest, a reluctance which may be a reflection of their comfort level with computers in general and TM systems specifically, an observation also made by Dillon and Fraser (2006, 79). Vendors should be aware that potential TM users may not have the technical knowledge to base their tool choice on features and functionalities, but may instead look to the potential benefits of TM usage when choosing to use a TM system or choosing a particular brand. I believe that vendors' promotional materials should equally reflect functionalities and the benefits that users may receive, and perhaps focus on learnability, ease of use, and training in order to better address issues more specific to new or potential users.

4.2.1.1 Situational considerations

4.2.1.1.1 Text characteristics and format

Discussions about purchasing or acquiring a TM system often associated the usefulness of a TM system with text characteristics and format (subject field, client, repetition, text type), and, while many linked usefulness to repetition, others indicated that the use of TM systems was also beneficial with non-repetitive texts, an observation also made by García (2006). Potential users may wish to investigate how TM systems

may be useful to translate non-repetitive texts as well as repetitive texts, and therefore also consider the effects of TM usage on their workflow and translation process. Current users may be able to expand their level of TM usage to include non-repetitive texts, perhaps especially if the system offers sub-segment matching or integrates EBMT. I believe that further investigation of the use of TM systems for non-repetitive texts is needed to determine the types of users who translate all texts with TM systems, why (i.e. the perceived and attained benefits), brand-specific relations to TM usage for non-repetitive texts, method-based relations (i.e. CSB versus SB approach), reasons for the appearance of this type of usage (i.e. as a result of research advances in EBMT or sub-segment matching), and the effects this type of usage has (or may have) on TM system adoption. Vendors should also research this area of TM usage as it may increase the marketability of their tool and allow them to promote it to new groups of users or types of users. Finally, I encourage trainers not only to address the use of TM systems for repetitive texts in training programs, but also to introduce students to other potential applications of TM systems.

An observation made in relation to the types of authors may help to explain why so few postings related to the use of hard-copy documents despite the significance of their role in non-adoption as identified in section 1.3.1.1. The use of electronic documents is a prerequisite to TM usage, and investigations of TM systems would take place after investigations of conversion methods. Thus, it can be assumed that potential users were already working primarily with electronic documents (or had already investigated means for conversion). This under-representation may also indicate that current TM users were performing the necessary steps for conversion in order to use TM systems, but not discussing this process or were discussing in another type of forum (e.g. forums related to OCR software or VRS). I believe these are the two factors contributing to the lack of discussion; however, in order to determine the current role of hard-copy document usage in non-adoption, further investigation is needed.

4.2.1.1.2 Number of users

I believe that the lack of discussion pertaining to networking capabilities of TM systems is an indicator of the type of authors; the authors appear to be primarily freelance

translators, and therefore not necessarily concerned with such issues. This consideration would most likely be more significant for agencies or organizations investigating TM systems, and therefore is not strongly represented in the postings.

4.2.1.1.3 Languages

Although the language of the TM interface was not discussed in the postings in the corpus, users should be aware that the availability of the interface in their first (or preferred) language could positively affect the learning curve. Therefore, I believe this should be an important consideration when choosing a TM system. Open-source software may be offered in a wider variety of languages and may meet the interface language needs of a user better than closed-source software. As noted by Bowker, McBride, and Marshman, SDL Trados “presently provides the user interface in four languages, while OmegaT [...] has been localized by volunteers into 20 languages” (2008, 21).

To date, no research has been conducted that indicates the extent to which the language of the interface (and supporting documentation) affects the learning curve; therefore, a deeper understanding of the relationship between the two would help translator trainers better address the needs of their students and provide further guidance for potential users when investigating TM systems. Furthermore, trainers should encourage students to consider the effect that the language of the interface, supporting documentation, and training materials may have on their learning curve.

Finally, vendors and developers should research language-specific issues in order to determine the areas (or languages) in which users’ needs are not being met and the effect of language support on brand choice.

4.2.1.2 Technical considerations

4.2.1.2.1 Computing requirements

Few discussions in the corpus pertained to processing speed and memory requirements. Two factors may have contributed to the lack of discussion surrounding these issues: authors seemed to understand that certain capabilities are necessary to properly run TEnts on personal computers, and the increasing processing speed and

memory of personal computers, especially in the past decade, may have reduced the number of issues related these requirements. Nevertheless, these issues are important as they directly affect the functioning of TM systems. I believe that it is important for users to be aware of the importance of meeting vendor-specified requirements, and that vendors should be aware of the importance of providing easily accessible, up-to-date, and detailed information about technical specifications.

The choice of TM system will depend on the user's OS, and, as such, the number of TM systems available may be limited for Linux and Macintosh OS users. While the number of Windows-based TM systems is significantly larger than that of Macintosh- or Linux-based systems, the contents of the OS-related postings suggest that there are many non-Windows users investigating TM systems. I encourage users to continue to vocalize their OS preferences in order to ensure that vendors are aware that cross-platform tools are needed and being used. Vendors should recognize that cross-platform capabilities may increase sales and research the potential for expanding adoption to users of other OSs. Further investigation into the OSs currently being used by translators may help to identify the demand for Macintosh- and Linux-compatible TM systems and to establish if users' current OS needs are being met with the cross-platform tools already on the market.

4.2.1.2.2 Open-source versus closed-source software

The corpus contained numerous discussions about open-source software; this may be a reflection of the proportion of freelance translators whose postings were included in the corpus, as freelance translators may be more concerned with price and may use a wider variety of OSs. Furthermore, freelance translators perhaps do not need the sophisticated quality assurance and project management features often included in more expensive TEnTs. While awareness seemed to be relatively high, I believe that users should investigate the availability of open-source tools (especially TM systems or TEnTs), and I encourage them to seriously consider these tools when choosing a TM system. There are open-source TM systems currently available, and enthusiasm for and use of these tools will most likely increase as the open-source movement develops. Indeed, this seems to be happening, as demonstrated, for example, by the 2008

conference hosted by the Localisation Research Centre in Ireland, which had the theme “Localisation 4 All”, with a heavy focus on open-source systems.⁶⁴ Researchers should investigate the uptake of open-source software applications by translators in order to determine how and why these programs are currently being used (or not used).

4.2.1.2.3 Word-processor or text-editor integration

The integration of TM systems into third-party word processors or text editors is currently being replaced by the incorporation of a proprietary text editor in many TEnTs (indicated by the release of Wordfast Professional and SDL Trados Studio 2009). While more experienced TM users may prefer proprietary text editors, all users should still investigate the need for additional software, and how integration of a word processor or text editor affects the ease of use, learnability, and user-friendliness of the display. Vendors and developers should remain in contact with their customers to determine attitudes towards the current state of integration and realize that users who are content with their current integrated systems may not easily migrate to a proprietary text editor. Researchers should investigate this area further in order to better determine current attitudes towards integration versus proprietary editors and identify concerns to allow for a smoother transition. Finally, trainers may need to adjust their current strategies to meet this new development.

4.2.1.2.4 Voice recognition software

The use of voice recognition software was discussed in the corpus. Users seemed to be aware that the use of VRS with a TM system may cause compatibility and conflict issues and were simply enquiring about other users’ experiences. Though not an issue of major significance for the use of TM systems, users do need to be aware of the problems they might encounter when using VRS with TM systems and investigate the specific compatibility of the two software applications.

⁶⁴ For more information on this conference, see <http://www.localisation.ie/resources/conferences/2008/index.htm>.

4.2.1.2.5 TM creation, storage, and usage methods

The corpus contained no discussions related to TM creation, storage, and usage methods (i.e. interactive creation and post-translation alignment, SB approach versus CSB approach, interactive translation and pretranslation). The lack of discussion related to creation methods may be a reflection of the type of users whose postings were included in the corpus: new or potential users may be unaware of the different methods for creation (and new translators may have few legacy documents), whereas experienced users have most likely already aligned their legacy translations and therefore create TMs through interactive creation.

I believe that the absence of discussion related to the SB versus CSB approach is primarily due to a lack of awareness. Furthermore, I expected that this issue would be more strongly represented in current research and literature, rather than by current practitioners, as the CSB approach is less common and less well known than the SB approach and discussions about the two approaches tend to be highly technical. Nevertheless, users should be aware of these different approaches (especially the SB approach and CSB approach) as one method may be more effective for their translation work. Researchers should continue to investigate the benefits and drawbacks in order to increase awareness and to aid translators when choosing a TM system. I believe that vendors of systems using the CSB approach should strongly promote their products and the specific benefits as potential users are most likely unaware of the CSB approach.

I am somewhat surprised by the lack of discussion pertaining to pretranslation, especially as the authors expressed strong opinions on other issues related to client requirements or demands (i.e. brand or format specifications, payment). The absence of any reference to pretranslation may simply be a result of the random selection method, an indication that it is currently not an issue of significance for TM users, or a reflection of the fact that pretranslation is largely determined by the client or agency and not the translators themselves. Further investigation of the use of pretranslation and translators' attitudes is necessary to determine its role and implications in the industry.

4.2.1.3 Brand-specific considerations

4.2.1.3.1 Trial or demo versions

New and potential users are strongly encouraged to investigate and use trial and demo versions as their use can provide the most effective method for determining whether or not a particular brand is the most appropriate for their situation. The choice of a TM system is personal and involves the consideration of numerous factors, and the use of a trial or demo version can help users to avoid purchasing a tool which later may be found to be unsuitable. I strongly believe that vendors should continue to provide and promote trial and demo versions in order to aid potential users in making a decision and should consider the type of trial or demo version offered (i.e. time-limited, size-limited, feature-limited), as demo versions with limited functionality may not allow the user to make an accurate evaluation of the functioning of the tool.

4.2.1.3.2 Price

The cost of investing in a TM system can be high; however, users should not be guided solely by price when choosing a TM system. Potential tools should be investigated and the decision should be based on numerous factors (including features, ease of use, and technical support): the most expensive tool is not always the ‘best’, and a lower purchase price does not necessarily indicate a lack of development or support. The choice to use a TM system is the choice to invest time (and usually money) in a product with expected returns, and these returns are not necessarily realized in the short term.

Vendors should note that price is a significant motivating factor for potential buyers, and as such, the benefits of TM usage (and ROI) should be adequately addressed in promotional materials in order to better guide new and potential users.⁶⁵ Inexperienced TM users will most likely be more attracted by the benefits offered by a tool, rather than the technical specifications of the tool’s features. However, it should be noted that ROI is largely dependent on the user’s particular context (e.g. type and volume

⁶⁵ SDL Trados does provide a free ROI calculator; however, its calculations are largely based on the amount of repetition among texts. This method of calculating ROI does not take the benefits of using TM systems for non-repetitive texts into consideration.

of work, familiarity with computers and TM systems); therefore, promotion of ROI could be misleading or exaggerated for some users.

Clients and agencies should also be aware of the potentially high price of many TEnTs and, in situations where particular brands or proprietary formats are required, consider options for helping the translator to make the initial investment (e.g. with the use of systems that allow translators to use the tool for free or advances on payment).

4.2.1.3.3 Updates or new releases

The availability and pricing policy for version updates and new releases were discussed frequently within the corpus. Users should be prepared for a delay in the release of patches and service packs (e.g. to address issues of compatibility with new OSs and other integrated software). Those looking to purchase a TM system should investigate the possibility of upcoming new releases and perhaps readjust time of purchase accordingly (i.e. wait until an upcoming version is released or until compatibility issues have been resolved). Furthermore, users should investigate the vendor's pricing policy for new releases, as this may affect the decision of which tool to purchase. Nevertheless, users need to realize that TM systems are not a one-time investment; they must be updated like any other software application, piece of technology, or other translation resources (e.g. word processors, dictionaries, computers).

Vendors should continue developing patches and service packs as quickly as possible and ensure that their customers are aware of the timeline for development. As many contributors were discontented with current pricing policies for new releases, I encourage vendors to investigate the attitudes of their customers and potentially adjust the product price or new release pricing policy (e.g. by slightly increasing the initial price or by implementing a set a fee for a specific period of time), as additional investment does not encourage those who are already hesitant about the price of a TM system to buy a product. The corpus discussions contained some support for Wordfast's three-year policy, and perhaps this model could be adopted and adapted by other vendors. Additionally, vendors could support older versions of TM systems for longer, enabling both clients/agencies and translators to receive the necessary technical support and access newer file formats.

4.2.2 Using a TM system

4.2.2.1 *Training*

Learning to use a tool properly involves the investment of time and money (either directly for paid training or indirectly in income lost while learning). Users should consider these investments when purchasing a tool and realize that their comfort level with TM systems and computers in general will affect their individual learning curve. The price of training was not commonly discussed in the corpus, perhaps indicating that the authors were more certain of this investment than of that in purchasing a tool, or that those who have purchased a tool are committed and willing to take the necessary steps to learn how to use it properly and efficiently. This recognition is important, because while discussion forums can be a source of help, they cannot provide adequate training; members of the two discussion boards are willing to help, but the user should first attempt to find the necessary information elsewhere. Users should also consider their preferred method of training in order to ensure that they will have access to training that suits their learning style and needs. No postings discussed training received from academic institutions or professional associations, despite the identification of these sources of training as valuable in the surveys examined. This may perhaps indicate that the contributing authors did not consider these sources of training to be valuable, that they simply were not considered to be sources of training, or that due to the regional nature of these types of training, they were not necessarily pertinent for users in other locations. I encourage TM users to offer feedback to the providers of the training (e.g. vendors, professional associations, universities or academic institutions), to allow them to better address the issues of importance to their audience.

Vendors should investigate the training needs of current and new users in order to provide the necessary means for users to learn how to use their tool. Furthermore, I encourage vendors to seek feedback about training formats and materials from their customers and address the areas for improvement identified.

Researchers and translator trainers are encouraged to continue investigating effective methods for the teaching of translation technologies. I believe that the functioning and features of TM systems are adequately addressed in current textbooks and other academic materials; however, these need to be expanded to include the issues

surrounding the use of TM systems (e.g. compatibility, client/agency relations, payment practices). Trainers are also encouraged to remain informed of current uses of and developments in TM systems and the issues surrounding their use in order to adequately and properly prepare future translators for the profession. In order to provide a balanced view of TM usage, trainers are encouraged to remain in contact with all stakeholders in the translation process and to explore new methods of incorporating multiple attitudes and viewpoints into training programs. While the time and budgetary constraints and learning curve involved in translation technology courses often do not allow for in-depth study of more than one brand of TM system, trainers are encouraged to inform their students about a wide variety of tools (e.g. cross-platform TM systems, open-source and closed-source systems, systems offering SB and CSB approaches, systems that use an integrated (third-party) word processor or text editor versus those with proprietary text editors, systems from all brackets of the price range) in order to increase awareness of the issues involved in choosing a TM system and to present the benefits and drawbacks of different types of TM systems to students. This does not need to be addressed with specific training on a wide variety of brands, but may be achieved with a relatively brief review of the tools and types of tools available on the market.

4.2.2.2 File formats

Issues relating to file formats (including issues of compatibility, vendor lock-in, and open standards) were the most discussed in the corpus: almost half of all postings contained discussions relating to these issues. Users should ensure their TM system can handle and produce the formats they commonly work with, investigate compatibility options (e.g. import and export capabilities), and reflect on how their choice of tool may affect future work (e.g. the jobs they may or may not be able to accept, the ease of working with various file types). Users should be aware of vendor lock-in (which can result, for example, from update pricing policies, client/agency requirements, and/or a lack of compatibility) and perhaps consider various tools that could produce the necessary format(s) to determine which one(s) they may find more pleasing to use. As identified by the corpus contents, SDL Trados is not the only option, and I strongly encourage users to research a number of tools (and all aspects of these tools) in order to

make the decision based on what suits their needs best. Users are also encouraged to support open standards, as they offer an effective and reliable method of exchanging data (although, like all methods for data exchange, perhaps are not entirely infallible).

I believe that vendors should continue to develop file format capabilities by increasing the number of formats that can be handled by a tool, creating filters to properly convert from one format to another, and addressing issues of formatting (e.g. font, styles, document layout). This is an important area for development needing continual attention. To combat vendor lock-in and compatibility issues, vendors are encouraged to support open standards and develop products which can read and produce open standard formats. I believe that this may increase customer satisfaction and would facilitate the process of switching from one TM system to another (based on their various strengths or client brand or format requirements), thereby further encouraging the use of multiple tools.

Researchers should continue investigating the formats used by translators in order to aid vendors and developers in determining the needs of TM users and identifying any current deficiencies in file format support. There has been little research conducted investigating users' cross-over among tools and current strategies for dealing with compatibility issues, though it is clear from the literature and the corpus contents that many translators do use multiple tools. Researchers should address these issues and also explore the use of open standards to determine when they are being used (or not being used) by vendors, TM users, and clients/agencies, and to determine how open standards can better meet the industry's needs.

Finally, trainers should ensure that training materials (i.e. textbooks) address the issues of compatibility and open standard formats and that students are introduced to these concepts in training courses.

4.2.2.3 Matches proposed by a TM system

The usefulness of the matches proposed by a TM system can depend on the size, source, and subject-field division of a TM. I recommend that users experiment with the size of their TMs in order to determine whether one large TM or smaller specific TMs better suit their needs. Additionally, users should experiment with the fuzzy match value to find the most appropriate threshold based on their editing versus translating

preferences. The choice and effects of fuzzy match values were not discussed by contributing authors, even though these issues were addressed in the literature. This absence may be an indication that the fuzzy match value is dictated by the client and is not the choice of the translator, or it may be due to the personal nature of fuzzy threshold preference. Additionally, the lack of discussion may be a reflection of the types of users whose postings were included in the corpus: new TM users may be unaware of the effect of the fuzzy threshold on TM usage, whereas experienced users may have already experimented with the fuzzy match value and found the threshold that suits them best. Users should also consider the usefulness of others' TMs and the amount of editing needed when using a TM from another source. Depending on the source of a TM and the amount of editing of the proposed translations that is required, these TMs may be valuable only as reference materials or may be extremely useful and provide helpful matches. Clients/agencies should ensure that their TMs are accurate, up-to-date, and of high quality, as this will affect the quality of matches proposed to (and subsequently used by) their translators. To date, relatively few examinations of these issues have been made, and researchers are encouraged to address these issues in future investigations. Trainers should address the strategies for TM division (i.e. one large TM versus smaller more specific TMs) in training courses and ensure that students are aware of the consequences of each strategy.

4.2.2.4 Technical issues

The corpus contained very few discussions related to customization of keyboard shortcuts and no discussions related to display customization. This may be due to the individual and personal nature of these customization options; each user has his or her own preferences. However, those who are interested in customization options should consider the importance of customization before choosing a TM system, as availability and level of customization differ from tool to tool. Specific customizability options do vary; nevertheless, the availability of and need for such options is important, as these options allow users to work comfortably with the tool. Vendors are encouraged to continue offering customization, or to investigate how their tool could be customized in order to better meet the needs of the users.

The availability and quality of technical support were the subject of numerous discussions in the corpus. I encourage new users to investigate vendors' reputations for technical support, the format of support provided (e.g. phone, e-mail, online), and the price of technical support before choosing a TM system. I also feel that current users should remain in contact with customer service departments, in order to ensure that current technical issues and solutions are made known to the vendor. While the members of discussion boards can provide invaluable help with technical issues, users should first research the problem (in knowledgebases, FAQs, previous postings on forums) before turning to their peers. Members of the two sites were very willing to help, but appreciated the use of forums as a complement to other sources rather than the sole source of technical support. After experiencing the sense of community on both ProZ.com and TranslatorsCafé.com, I strongly recommend that users continue to support each other and share advice and tips with their fellow forum members.

Vendors should maintain accurate and current knowledgebases and FAQs to guide users when encountering technical problems. Furthermore, I encourage vendors to investigate their own reputations for technical support in order to address the issues as identified by their customers. As seen through the analysis of the corpus contents, vendors do have reputations for customer support, and discussion boards could provide valuable insight into the opinions of the users of a particular tool brand.

4.2.2.5 Procedural and textual effects of TM usage

Numerous postings in the corpus discussed the effects of TM usage on speed; these discussions were closely related to textual characteristics (e.g. text type, repetition, subject field), ROI, and training. Based on the content of these postings, authors seemed to recognize that the speed of translation is directly affected by many other issues (such as those mentioned above, as well as familiarity with computers and previous experience using TM systems) and that an increase in productivity is not always immediate. I believe that potential users should consider all aspects of their translation process and context when deciding to use a TM system and that it is especially important to recognize that the use of TM systems (and subsequent realization of the benefits) may require a long-term investment.

While the quality of translations produced using TM systems was little discussed in the corpus, both positive and negative effects of TM use were mentioned. Authors seemed to recognize that matches found in a TM (whether it was personally created or supplied by a client/agency) require careful revision and evaluation in order to ensure they are appropriate for the context. Users need to recognize, in the words of one author, that a “*tool [is] nothing more than that - a tool*” (TC48-08) and quality is not a result of TM use, but of the translator’s abilities. Clients and agencies should ensure the TMs sent to their translators are accurate, up-to-date, and of high quality in order for them to be of assistance to the translators. Trainers need to maintain a focus on critical use of tools and not allow trainees to fall into the trap of blindly accepting whatever solution is proposed by a TM system.

The corpus discussions referenced procedural benefits of TM usage (e.g. to maintain the formatting in the text, to reduce the amount of toggling between windows, and to present the text segment-by-segment in order to ensure nothing is missed when translating). These types of benefits are currently under-represented in the literature; this disparity may be due to the fact that the procedural benefits of TM usage become more apparent when translating non-repetitive texts (i.e. when the system is not providing matches from a TM), an application of TM systems that is seldom discussed in the literature. Further investigation is needed to determine the procedural benefits of TM usage and the contexts in which these benefits may be most apparent. Conversely, the corpus contained no discussions related to the procedural drawbacks of database maintenance and alignment verification, perhaps indicating that although authors may be aware of the procedural drawbacks, they simply consider them to be unavoidable and necessary (i.e. inherent in TM usage).

4.2.2.6 Client relations

Almost 20% of the postings in the corpus contained discussions related to client/agency brand and format requirements, indicating that this is an important issue for TM users. Users are encouraged to investigate methods for compatibility to allow for the use of their preferred brands when the requirements of the client or agency do not. Clients and agencies should communicate to their translators the reasons behind their tool

choice and perhaps offer concessions when the requirement cannot be directly fulfilled (e.g. use of open standards, help in making the initial investment in the required tool). Researchers should explore clients' and agencies' attitudes and perceptions of the use of TM systems in order to determine their motivations and to help bridge the gap between translators and clients/agencies.

The issues of payment for matches found in a TM was also often discussed in the corpus. If required to use a TM for their work, users should establish rates which reflect the amount of time needed to evaluate and edit matches. As the use of TM systems is relatively recent, there are currently no industry-wide practices, and rates are determined on a contract-to-contract basis; therefore, users are able to set rates which ensure that they are compensated for all of their work. However, translators may feel "undercut" by competitors who set lower rates. In order to effect the change that some TM users seem to want, translators need to discuss the issues among themselves, recognize the impact of their choices and not "undercut" competitors, and work together to establish industry standards. Researchers and professional associations are encouraged to explore the state of payment practices in the industry in order to establish current rates of pay for matches found in TMs. Furthermore, future research should investigate the amount of time involved in editing exact and fuzzy matches (versus translating the entire segment). I believe that clients and agencies need to realize the time investment needed when dealing with exact and fuzzy matches and negotiate rates that reflect this.

I encourage trainers to ensure that training materials (i.e. textbooks) address the issues of client brand or format requirements and payment practices for matches found in a TM, and that students are introduced to these issues in training courses.

Finally, I strongly encourage all stakeholders to participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs.

I can offer no possible explanations for the lack of discussion relating to ownership of TMs (aside from being a result of the random method used for thread selection). I hesitate to suggest that this may indicate that ownership is not as significant of an issue to TM users as described in the literature; further research related to

ownership is clearly needed to help determine the importance of the issue for both TM users and clients/agencies.

The corpus contained very few references to the concept of buying/selling/trading TMs, perhaps indicating that translators are uninterested in or unaware of the potential benefits of a TM marketplace. The findings of the LISA 2002 and 2004 surveys also indicated a lack of interest due to legal complications; however, the appearance of companies and organizations focused on the exchange of TMs indicates that perhaps these challenges have been at least partially overcome (see examples of the TDA and TM Marketplace in section 1.4.6.4). Further research is needed to determine when and by whom TMs are being bought/sold/traded and why translators may not necessarily be discussing the benefits or drawbacks of these exchanges.

4.3 Evaluation of the methodology

In the *Scope and Limitations* in the introduction (section 0.3), specific criteria were established for evaluating possible primary sources. The primary source needed to contain translators' *unprompted* thoughts and opinions relating to TM usage. The source needed to provide *sufficient recent TM-related data in English*. The primary source was not to focus on specific *geographical locations* or *language pairs*. Furthermore, the data was to be accessible by means *other than keyword searches or brand-name based collection*. Finally, only *data related to TM systems*, and not to other TEnT components, was analyzed.

TranslatorsCafé.com and ProZ.com satisfied all of these criteria. The availability of postings written by a wide variety of authors in various locations working in many different languages allowed for the inclusion of a diverse range of opinions from translators working in numerous contexts.

The creation of a sample corpus allowed me to determine the type and amount of information related to TM systems and consequently adjust the start of my date range (from January 1, 2005 to January 1, 2007) in order to ensure a manageable number of postings was collected for classification. The use of postings written after December 31, 2006 allowed for the collection of recent data. In fact, the lack of discussion related to language support (for right-to-left languages or double-byte characters) may be a

reflection of the currency of the postings, as many TM systems have become Unicode compliant in recent years, likely reducing problems related to language support.

The scope of the data was restricted to postings specifically related to TM systems, and this may have eliminated discussions of issues closely related to the use of TM systems. For example, the corpus contained no discussions of legacy file formats, which may be attributed to the exclusion of discussions focusing on bitext aligners, and relatively little discussion of the conversion of hard-copy documents to electronic format, which may be attributed to the exclusion of discussions focusing on VRS or OCR software.

As the corpus contained only postings in English, certain language-specific issues may have been discussed less frequently; for example, the corpus contained no discussions relating to the language of the interface of TM systems and relatively few postings pertaining to language-pair considerations. This does not necessarily indicate that these issues are unproblematic or irrelevant, simply that the users who may encounter these difficulties are not discussing them in the threads included in the corpus; other language-specific or non-English forums may contain discussions related to these topics. Users encountering language-specific difficulties (such as those encountered when working with right-to-left languages or double-byte characters) may also be discussing these issues in their respective language-specific forums (e.g. Arabic, Hebrew, Japanese, Korean). Furthermore, as authors of the postings used in the analysis are presumably comfortable working and/or communicating in English and the majority of TM systems have English-language interfaces, this issue may not be particularly problematic for them, or may not play an important role when choosing a TM system. While the lack of discussion pertaining to these issues may indicate that users' needs are currently being met, further investigation of other sources would be needed to indentify the extent to which the language of the interface, and right-to-left language and double-byte character support are being discussed and to determine whether or not the systems currently available have adequately addressed the issue. This type of investigation is beyond the scope of this project given the associated language requirements.

In order to allow for the collection of non-brand-specific postings without primarily using keyword searches, threads were randomly selected for inclusion in the

corpus. This random selection process may have contributed to the omission or underrepresentation of discussions related to certain issues, especially in the case of the threads taken from ProZ.com, as only every twelfth thread was selected. However, I believe that the use of a random method for thread identification provided an accurate overview of the issues being discussed and allowed for the systematic yet neutral identification of threads considering the time constraints of this project. A method based on brand-specific identification or based on keyword searches could affect the type of opinions within the postings or the types of issues being discussed.

Overall, I feel that the implementation of the methodology developed for this project was successful, although the forum and thread selection processes, the creation of the corpus, and the classification of the postings were very labour-intensive. It was impossible to anticipate the amount and type of data on the discussion boards; therefore, I needed to ensure that my methodology could be refined and adjusted throughout the process. The development and implementation of the methodology was a very complex multi-step process; however, the complexity of the methodology seemed unavoidable in order to ensure the representativity of the corpus. The use discussion forum postings for this project has aided in identifying specific challenges that may arise when analyzing this type of primary source and in developing strategies for addressing these challenges. Finally, the methodology provides general principles for selecting and managing postings, and these could be applied to future examinations of discussion board postings.

The classification of the data was equally demanding given the challenges associated with the annotation and interpretation of open-ended responses (as opposed to the closed-ended responses of surveys). However, the flexibility of the classification system allowed me to modify the categories at any point the process in order to organize the data in an informative and logical way. The classification system encompasses numerous TM-related issues and could be reapplied by any of the stakeholders to guide future evaluations of TM systems.

The choice of discussion boards as a primary source and the subsequent classification of the postings have provided me with the opportunity to observe how various points of view complement one another and contribute to the overall view of TM systems. By comparing and contrasting these different viewpoints, I was able to identify

areas of conflict among users and various stakeholders and issues which are currently under-represented in the literature of the field.

Although the methodology described in Chapter 2 allowed for the collection of postings from ProZ.com and TranslatorsCafé.com according to the criteria identified in the *Scope and Limitations* (section 0.3), some adaptations to this methodology may be needed for use with other forums, or even for reuse with the two discussion boards chosen for this project. The content and structure of forums is continually changing, and a previously developed methodology may not provide the expected results when implemented in a new context. In fact, even though the time period between data identification and writing was short (less than six months), the “CAT and other Software for Translators” forum on TranslatorsCafé.com has since undergone extensive reorganization according to the brand names of the tools discussed, and TM-related threads are now primarily created and posted on in brand-specific forums.

In spite of the specific challenges associated with a reimplementations of the methodology used for this project, further investigation of translators’ unprompted and spontaneous opinions may prove to be interesting in numerous other contexts. Future analyses of postings on the two discussion boards used for this project, or on other translation-focused boards with regard to other CAT tools or MT systems, specific TM-related issues, particular brands, or specific time periods (to track the evolution of tools in general or of a specific brand, or to observe discussions sparked by industry news, events, or announcements), specific languages or regions (for comparative analyses), or additional investigation of the issues addressed in this project (in order to highlight developments of TM systems or progress within the field) may provide a deeper understanding of translators’ attitudes in these specific contexts. Finally, the findings of this analysis could be used to develop more specific, relevant, or focused surveys in order to address TM system uptake and the issues surrounding use, and subsequently expand and broaden the current body of knowledge.

4.4 Concluding remarks

TM systems are perhaps the most aggressively marketed and widely used CAT tools in the industry. Although TM systems have been commercially available for some

time, research into the use of these systems and the surrounding issues is more valuable than ever. Technology continually advances and develops, and these changes impact all stakeholders involved. While it may seem like a battle to keep up with current trends and developments, it is imperative that stakeholders establish and maintain open lines of communication in order to overcome the challenges related to TM use, as I believe that this is the first step towards resolving some of the conflicts observed in the corpus postings. I hope that this thesis assists in increasing awareness of translators' attitudes towards TM systems, helps to reveal the importance of the issues for different stakeholders, and contributes to the communication among stakeholders needed to develop tools that benefit translators and their clients.

References

- Allied Business Intelligence (ABI). Language Translation, Localization and Globalization: World Market Forecasts, Industry Drivers, and eSolutions. Oyster Bay, NY: Allied Business Intelligence, 2002.
- aquarius.net. "About Aquarius." 2009a. 8 May 2009
<http://www.aquarius.net/index.cfm?aq=demo_about>.
- . "Answers." 2009b. 8 May 2009 <<http://www.aquarius.net/answers/>>.
- Association of Translators and Interpreters of Ontario (ATIO) Salaried Translators Committee. Results of the 2007 Survey of Salaried Translators. Association of Translators and Interpreters of Ontario, 2007. 11 March 2009
<http://www.atio.on.ca/Membership/Sal_Survey/Sal_Tran_Srvy_Rslds.asp>.
- ATRIL. "Déjà Vu / Trados Compatibility Guide." 1999. 2 June 2009
<<http://www.atril.com/docs/DV-Trados.doc>>.
- . "Multiple File Format Support: Everything Under One Roof." 2009. 28 May 2009
<<http://www.atril.com/overview/formats.asp>>.
- Austermühl, Frank. Electronic Tools for Translators. Manchester, United Kingdom: St. Jerome Publishing, 2001.
- Automatic Language Processing Advisory Committee (ALPAC). Language and Machines: Computers in Translation and Linguistics. Washington, D.C.: National Research Council, 1966.
- Bowker, Lynne. Computer-Aided Translation Technology: A Practical Introduction. Ottawa: University of Ottawa Press, 2002.
- . "Investigating 'Reversible' Translation Resources: Are they Equally Useful in both Directions?" Speaking in Tongues: Language Across Contexts and Users. Ed. Luís Pérez-González. Valencia: Universitat de Valencia, 2003. 201-224.
- . "What does it Take to Work in the Translation Profession in Canada in the 21st Century? Exploring a Database of Job Advertisements." META 94.4 (2004): 960-72.
- . "Translation Memory and "Text"." Lexicography, Terminology and Translation: Text-Based Studies in Honour of Ingrid Meyer. Ed. Lynne Bowker. University of Ottawa Press, 2006. 175-187.
- Bowker, Lynne, and Carolina Herrera. "A New Internet-Based Communicative Setting: Exploring its Impact on Language for Special Purposes." Lexicographica 20 (2004): 33-5.

- Bowker, Lynne, Cheryl McBride, and Elizabeth Marshman. "Getting More than You Paid for? Considerations in Integrating Free and Low-Cost Technologies into Translator Training Programs." Revista Electrónica de Didáctica de la Traducción y la Interpretación 1 (2008): 26-47. 16 May 2009 <http://www.reedit.uma.es/doc/Bowker_et_al.pdf>.
- Canadian Translation Industry Sectoral Committee (CTISC). Survey of the Canadian Translation Industry: Human Resources and Export Development Strategy. Ottawa/Montreal: 1999. 8 April 2008 <<http://www.uottawa.ca/associations/csict/princi-e.htm>>.
- Clegg Smith, Katherine M. "'Electronic Eavesdropping': The Ethical Issues Involved in Conducting a Virtual Ethnography." Online Social Research: Methods, Issues, & Ethics Ed. Mark D. Johns, Shing-Ling Sarina Chen, and G. Jon Hall. New York: Peter Lang, 2004. 223-238.
- Computer Language Company Inc. "Operating System." PCMAG.COM. 2009. 20 May 2009 <<http://www.pcmag.com/encyclopedia/>>.
- de Vries, Arjen-Sjoerd. "Getting Full Or Fuzzy? The Payment Issue with Full Matches Generated by Translation Memory Systems." Language International 14.3 (2002): 44-7.
- DePalma, Donald A. SDL-TRADOS: Language Service Provider Reaction SDL's Purchase of TRADOS. Andover, MA: Globalization & Localization Association, 2005. 3 June 2009 <http://www.gala-global.org/en/resources/downloads/2005Q3_GALA-SDL-TradosReport.pdf>.
- Dillon, Sarah, and Janet Fraser. "Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption." Machine Translation 20.2 (2006): 67-79.
- English Spanish Translator Org. "English Spanish Translator Org: Translation Forum - Translators Forums." 2006a. 8 May 2009 <<http://www.english-spanish-translator.org>>.
- . "English Spanish Translator Org: Translation Forum - Translators Forums: Members List." 2006b. 8 May 2009 <<http://www.english-spanish-translator.org/members/list/index25.html>>.
- European Commission Joint Research Centre. "The DGT Multilingual Translation Memory of the Acquis Communautaire: DGT-TM." 2008. 5 August 2009 <<http://langtech.jrc.it/DGT-TM.html#Download>>.
- Eysenbach, Gunther, and James E. Till. "Ethical Issues in Qualitative Research on Internet Communities." British Medical Journal 323 (2001): 1103-5.

- Fifer, Matthew. "The Fuzzy Factor: An Empirical Investigation of Fuzzy Matching in the Context of Translation Memory Systems." MA thesis. University of Ottawa, 2007.
- Fulford, Heather, and Joaquin Granell-Zafra. "Translation and Technology: A Study of UK Freelance Translators." The Journal of Specialised Translation 4 (July) (2005): 2-17. 6 November 2008
<http://www.jostrans.org/issue04/art_fulford_zafra.pdf>.
- García, Ignacio. "Standard Bearers: TM Brand Profiles at Lantra-L." Translation Journal 7.4 (2003). 1 April 2009 <<http://accurapid.com/journal/26tm.htm>>.
- . "Translators on Translation Memories: A Blessing Or a Curse?" Translation Technology and its Teaching (with Much Mention of Localization). Ed. A. Pym, A. Perecrestenko, and B. Starink. Universitat Rovira i Virgili, 2006. 97-105. 1 April 2009 <http://isg.urv.es/library/papers/Garcia_Translators.pdf>.
- Gauthier, François. 2006 Survey on Rates and Salaries. Trans. James Cookson. Ordre des traducteurs, terminologues et interprètes agréés du Québec, 2006. [Available to members only.].
- GlobalNet Services. "Global Langues - Forum List - Forum Centrale." 2009. 8 May 2009 <http://www.centrale-de-la-traduction.com/component/option,com_fireboard/Itemid,242/lang,en/>.
- Gow, Francie. "Metrics for Evaluating Translation Memory Software." MA thesis. University of Ottawa, 2003.
- . "You must Remember this: The Copyright Conundrum of 'Translation Memory' Databases." Canadian Journal of Law and Technology 6.3 (2007): 175-92.
- Harris, Brian. "Bi-Text, a New Concept in Translation Theory." Language Monthly 54 (1988): 8-10.
- Heartsome. "Heartsome Software Overview." 2008. 20 May 2009 <<http://www.heartsome.net/EN/products.html>>.
- Herring, Susan. "Linguistic and Critical Analysis of Computer-Mediated Communication: Some Ethical and Scholarly Considerations." The Information Society 12 (1996): 153-68.
- Hesseldahl, Arik. "Apple's Growing Bite of the Market." Business Week 2006. 20 May 2009 <http://www.businessweek.com/technology/content/jun2006/tc20060615_080175.htm>.

- Heyn, Matthias. "Translation Memories: Insights and Prospects." Unity in Diversity? Current Trends in Translation Studies. Ed. Lynne Bowker, et al. Manchester, United Kingdom: St. Jerome Publishing, 1998. 123-136.
- Höcker, Mary. Translation Memory Survey 2003. eContent Localization Resources for Translator Training, 2003. 6 March 2009
<http://ecolore.leeds.ac.uk/downloads/2003.05_bdue_survey_analysis.doc>.
- Hutchins, John. "The Origins of the Translator's Workstation." Machine Translation 13.4 (1998): 287-307.
- Kay, Martin. "The Proper Place of Men and Machines in Language Translation." Machine Translation 12 (1997): 3-23. [Originally published in 1980 as a Xerox PARC working paper.]
- Lagoudaki, Elina. Translation Memories Survey 2006. Imperial College London, 2006. 6 March 2009. <<http://www3.imperial.ac.uk/pls/portallive/docs/1/7307707.PDF>>.
- L'Homme, Marie-Claude. Initiation à La Traductique. 2nd ed. Montreal: Linguattech Éditeur, 2008.
- Linux Information Project (LINFO). "GUI Definition." 2004. 14 July 2009
<<http://www.lininfo.org/gui.html>>.
- . "Standards Definition." 2006a. 5 August 2009
<<http://www.lininfo.org/standards.html>>.
- . "Vendor Lock-in Definition." 2006b. 2 June 2009
<http://www.lininfo.org/vendor_lockin.html>.
- Localization Industry Standards Association (LISA). "Segmentation Rules eXchange (SRX)." 2008a. 26 May 2009 <<http://www.lisa.org/Segmentation-Rules-e.40.0.html>>.
- . "Translation Memory eXchange (TMX)." 2008b. 26 May 2009
<<http://www.lisa.org/Translation-Memory-e.34.0.html>>.
- Lommel, Arle. LISA 2002 Translation Memory Survey: Translation Memory and Translation Memory Standards. Localization Industry Standards Association, 2002.
- . LISA 2004 Translation Memory Survey: Translation Memory and Translation Memory Standards. Ed. Rebecca Ray. Localization Industry Standards Association, 2004.

- Macklovitch, Elliott, and Graham Russell. "What's Been Forgotten in Translation Memory." *Envisioning Machine Translation in the Information Future: Proceedings of the 4th Conference of the Association for Machine Translation in the Americas*. Cuernavaca, Mexico, 2000.
- McInnis, Nancy, and Maha Takla. 2005 Survey of Independent Translators. Association of Translators and Interpreters of Ontario, 2005. 22 October 2008
<http://www.atio.on.ca/Membership/Ind_Survey/Survey05_Intro.html>.
- Melby, Alan K. "Translators and Machines - Can They Cooperate?" META 26 (1981): 23-34.
- Merkel, Magnus. "Consistency and Variation in Technical Translation: A Study of Translators' Attitudes." Unity in Diversity? Current Trends in Translation Studies. Ed. Lynne Bowker, et al. Manchester, United Kingdom: St. Jerome Publishing, 1998. 137-149.
- Mogensen, Else. "Orwellian Linguistics." Language International 12.5 (2000): 28-31.
- National Library of Australia. "Australian Subject Gateway Forums: Vortal." 2007. 8 May 2009 <<http://www.nla.gov.au/initiatives/sg/servicetypes.html#vortal>>.
- Niska, Helge. "Lantra-L FAQ: Users' Guide." Lantra-L. 7 March 2009
<<http://www.geocities.com.proxy.bib.uottawa.ca/Athens/7110/ltr2.htm>>.
- O'Brien, Sharon. "Practical Experience of Computer-Aided Translation Tools in the Software Localization Industry." Unity in Diversity? Current Trends in Translation Studies. Ed. Lynne Bowker, et al. Manchester, United Kingdom: St. Jerome Publishing, 1998. 115-122.
- . "Eye-Tracking and Translation Memory Matches." Perspectives: Studies in Translatology 14.3 (2006): 185-205.
- OmegaT. "Other OmegaT Resources." 2009a. 28 May 2009
<<http://www.omegat.org/en/resources.html>>.
- . "Requirements." 2009b. 20 May 2009
<<http://www.omegat.org/en/requirements.html>>.
- Open Source Initiative. "The Open Source Definition." 2009. 15 July 2009
<<http://www.opensource.org/docs/osd>>.
- ProZ.com. "About ProZ.com: Site Rules." 2009a. 8 May 2009
<<http://www.proz.com/?sp=siterules&mode=show&category=forum>>.
- . "Frequently Asked Questions." 2009b. 10 May 2009
<<http://www.proz.com/faq/3097#3097>>.

- . "ProZ.com Privacy Statement." 2009c. 10 May 2009
<<http://www.proz.com/?sp=info/privacy>>.
- . "Translation Industry Discussion Forums." 2009d. 8 May 2009
<<http://www.proz.com/forum>>.
- Quah, Chiew Kin. Translation and Technology. Basingstoke, United Kingdom: Palgrave Macmillan, 2004.
- Rode, Tony. "Translation Memory: Friend Or Foe?" International Journal for Language and Documentation April (2000): 12-3.
- SDL. "SDL Knowledgebase: Glossary." 2009. 26 May 2009
<<http://talisma.sdl.com/display/2n/index1.aspx?tab=glossary&r=0.4980232>>.
- Smith, Jeffrey S. "Why Customers are Flocking to Linux." IBM. 2005. 20 May 2009
<<http://www-01.ibm.com/software/info/features/feb152005/>>.
- Somers, Harold. "Translation Memory Systems." Computers and Translation: A Translator's Guide. Ed. Harold Somers. Amsterdam: John Benjamins Publishing Company, 2003. 31-47.
- TAUS Data Association. "Join us today!" 2009a. 4 June 2009
<<http://www.tausdata.org/>>.
- . "Member Directory." 2009b. 4 June 2009
<<http://www.tausdata.org/index.php/members/member-directory>>.
- TERMIUM. "Operating System." Translation Bureau, Government of Canada. 2006a. 14 July 2009 <<http://www.btb.termiumplus.gc.ca>>.
- . "Text Editor." Translation Bureau, Government of Canada. 2006b. 15 July 2009
<<http://www.btb.termiumplus.gc.ca>>.
- TM Marketplace. "TM Marketplace White Paper - Sharing Translation Memory Data Aligned from Third-Party Documents: Legal Considerations." 2009. 4 June 2009
<<http://www.tnmarketplace.com/whitepapers/align.pdf>>.
- Topping, Suzanne. "Sharing Translation Database Information." MultiLingual. July/August (2000). 1 August 2009
<<http://www.multilingual.com/articleDetail.php?id=1105>>.
- TranslatorsCafé.com. "About TranslatorsCafé.com." 2009a. 8 May 2009
<<http://www.translatorscafe.com/cafe/History.asp>>.
- . "TranslatorsCafé Help: Forums: Rules of Conduct." 2009b. 8 May 2009
<<http://www.translatorscafe.com/tc/terms/EN/help/?ctg=C&HelpID=114>>.

- . "TranslatorsCafé.com Forums: Category / Forum Listings." 2009c. 8 May 2009
<<http://www.translatorscafe.com/cafe/MegaBBS/category-view.asp?showall=true>>.
- Wallis, Julian. "Interactive Translation vs Pre-translation in the Context of Translation Memory Systems: Investigating the Effect of Translation Method on Productivity, Quality and Translator Satisfaction." MA thesis. University of Ottawa, 2006.
- Wheatley, Alan. A Major Breakthrough for Translator Training. eContent Localization Resources for Translator Training, 2003. 6 March 2009
<<http://iti.org.uk/uploadedFiles/surveys/eCoLoRe%20results.pdf>>.
- Yahoo!. "What are the options for each of my Group subscriptions?" 2009a. 8 May 2009
<<http://help.yahoo.com/l/ca/yahoo/groups/messages/messages-01.html>>.
- . "Yahoo! Groups- Join or Create Groups, Clubs, Forums & Communities." 2009b. 8 May 2009 <<http://ca.groups.yahoo.com>>.
- . "Yahoo! Groups Search Results – 'translation memory'." 2009c. 8 May 2009
<<http://groups.yahoo.com/search?query=%22translation+memory%22>>.

Appendix A

Hierarchical Listing of Issues and Cross-References

Topic	Section in Literature Review (Chapter 1)	Section in Corpus Analysis (Chapter 3)	Section in Conclusion
<i>Purchasing or Acquiring a TM System</i>	1.3	3.2.1	4.2.1
<i>Situational Considerations</i>	1.3.1	3.2.1.1	4.2.1.1
text characteristics and format	1.3.1.1	3.2.1.1.1	4.2.1.1.1
number of users	1.3.1.2	3.2.1.1.2	4.2.1.1.2
languages	1.3.1.3	3.2.1.1.3	4.2.1.1.3
language pairs	1.3.1.3.1	3.2.1.1.3.1	
language of the interface	1.3.1.3.2	3.2.1.1.3.2	
<i>Technical Considerations</i>	1.3.2	3.2.1.2	4.2.1.2
computing requirements	1.3.2.1	3.2.1.2.1	4.2.1.2.1
processing speed and memory	1.3.2.1.1	3.2.1.2.1.1	
operating system	1.3.2.1.2	3.2.1.2.1.2	
open-source versus closed-source software	1.3.2.2	3.2.1.2.2	4.2.1.2.2
word-processor or text-editor integration	1.3.2.3	3.2.1.2.3	4.2.1.2.3
voice recognition software	1.3.2.4	3.2.1.2.4	4.2.1.2.4
TM creation, storage, and usage methods	1.3.2.5	3.2.1.2.5	4.2.1.2.5
SB or CSB approaches	1.3.2.5.1		
interactive creation or post-translation alignment	1.3.2.5.2		
pretranslation and interactive translation modes	1.3.2.5.3		
<i>Brand-Specific Considerations</i>	1.3.3	3.2.1.3	4.2.1.3
trial or demo versions	1.3.3.1	3.2.1.3.1	4.2.1.3.1
price	1.3.3.2	3.2.1.3.2	4.2.1.3.2
updates or new releases	1.3.3.3	3.2.1.3.3	4.2.1.3.3

Topic	Section in Literature Review (Chapter 1)	Section in Corpus Analysis (Chapter 3)	Section in Conclusion
<i>Using a TM System</i>	1.4	3.2.2	4.2.2
<i>Training</i>	1.4.1	3.2.2.1	4.2.2.1
time	1.4.1.1	3.2.2.1.1	
availability	1.4.1.2	3.2.2.1.2	
price	1.4.1.3	3.2.2.1.3	
<i>File Formats</i>	1.4.2	3.2.2.2	4.2.2.2
input and output files	1.4.2.1	3.2.2.2.1	
source and target file formats	1.4.2.1.1	3.2.2.2.1.1	
legacy file formats	1.4.2.1.2	3.2.2.2.1.2	
files created by the TM system	1.4.2.1.3	3.2.2.2.1.3	
TM formats	1.4.2.1.4	3.2.2.2.1.4	
compatibility among TM systems	1.4.2.2	3.2.2.2.2	
vendor lock-in	1.4.2.3	3.2.2.2.3	
open standard formats	1.4.2.4	3.2.2.2.4	
<i>Matches Proposed by a TM System</i>	1.4.3	3.2.2.3	4.2.2.3
types of matches	1.4.3.1	3.2.2.3.1	
usefulness of matches	1.4.3.2	3.2.2.3.2	
<i>Technical Issues</i>	1.4.4	3.2.2.4	4.2.2.4
customizability of TM systems	1.4.4.1	3.2.2.4.1	
technical support	1.4.4.2	3.2.2.4.2	
<i>Procedural and Textual Effects of TM Usage</i>	1.4.5	3.2.2.5	4.2.2.5
speed	1.4.5.1	3.2.2.5.1	
quality	1.4.5.2	3.2.2.5.2	
modifications to the translation process	n/a	3.2.2.5.3	
<i>Client Relations</i>	1.4.6	3.2.2.6	4.2.2.6
brand or format requirements	1.4.6.1	3.2.2.6.1	
payment	1.4.6.2	3.2.2.6.2	

ownership		1.4.6.3	3.2.2.6.3	
buying/selling/trading TMs		1.4.6.4	3.2.2.6.4	
<i>Other</i>				
non-adoption attitudes		1.6.3.3	3.2.2.7	
vendor promotional materials		1.5	3.2.3	

Appendix B

Bibliography of Vendor Promotional Materials

Across Systems Incorporated. Advertisement. MultiLingual. April/May 2009: 2.

---. “Across – Across Language Server – No. 1 Independent Technology for the Linguistic Supply Chain.” 29 May 2009.
<<http://www.across.net/en/index.aspx>>.

Advanced International Translations. “Advanced Software for Translators and Translation Agencies – Translation Memory Software.” 2009. 29 May 2009.
<http://www.translation3000.com/Translators_Software/Translation_Memory.html>.

Alchemy Software Development. “Products – Alchemy CATALYST 8.0.” 29 May 2009. <http://www.alchemysoftware.ie/products/alchemy_catalyst.html>.

---. Advertisement. MultiLingual. October/November 2007: 9.

Anaphraseus. “Anaphraseus.” 29 May 2009. <<http://anaphraseus.sourceforge.net/>>.

ATRIL. “Déjà Vu – Translation Memory and Productivity System.” 29 May 2009.
<<http://www.atril.com/>>.

CollabNet Incorporated. “Open Language Tools.” 2007. 29 May 2009. <<https://open-language-tools.dev.java.net/>>.

Heartsome Technologies Limited. “Heartsome New Generation Language Technology.” 2008. 29 May 2009. <<http://www.heartsome.net/EN/home.html>>.

---. Advertisement. MultiLingual. October/November 2008: 51.

Idiom Technologies Incorporated. Advertisement. MultiLingual. October/November 2007: 83.

JiveFusion. “Products – The JiveFusion Suite.” 2009. 29 May 2009.
<<http://www.jivefusiontech.com/products.php>>.

Kilgray Translation Technologies. Advertisement. MultiLingual. April/May 2009: G11.

---. “Benefits – Kilgray Translation Technologies.” 2009. 29 May 2009.
<<http://en.kilgray.com/?q=node/products/memoq/benefits>>.

Lingotek. “Translation Software – Lingotek.” 2009. 29 May 2009.
<<http://www.lingotek.com/EN/lingopoint.php>>.

---. Advertisement. MultiLingual. April/May 2009: 14.

Lingua et Machina. "Similis." 29 May 2009.
<http://www.similis.org/linguaetmachina.www/index.php?afficher=10&info=Similis>>.

Maxprograms. "Swordfish Translation Editor." 2009. 29 May 2009.
<http://www.maxprograms.com/products/swordfish.html>>.

MetaTaxis. "Why MetaTaxis?" 29 May 2009. <http://www.metataxis.com/why.htm>>.

MultiCorpora R&D Incorporated. "MultiTrans Automatic Translation Tools." 2009. 29 May 2009. <http://www.multicorpora.com/products/multitrans/>>.

---. Advertisement. MultiLingual. October/November 2007: 71.

MultiLing. "Fortis – Translation Memory Tool – Complete Translation System." 2009. 29 May 2009. <http://www.multiling.com/Fortis.aspx>>.

OmegaT. "OmegaT – The Free Translation Memory Tool." 29 May 2009.
<http://www.omegat.org/en/omegat.html>>.

Schaudin. Advertisement. MultiLingual. October/November 2007: 77.

---. "RC-WinTrans X8 Products." 31 January 2009. 29 May 2009.
http://www.schaudin.com:80/web/products_v8.aspx>.

SDL Trados Technologies. Advertisement. MultiLingual. October/November 2007: 84.

SDL Translationzone. "SDL Trados Suite 2007." 29 May 2009.
<http://www.translationzone.com/en/Products/sdl-trados-freelance/sdl-trados-2007-suite.asp>>.

Sisulizer Limited & Company KG. "Sisulizer – Visual Software Localization Tool." 2009. 29 May 2009. <http://www.sisulizer.com/>>.

STAR Group America, LLC. Advertisement. MultiLingual. October/November 2008: 3.

---. "Transit – The Translation Memory System." 29 May 2009. <http://www.star-group.net/star-www/description/transit/star-group/eng/star.html>>.

Terminotix Incorporated. Advertisement. MultiLingual. October/November 2008: 72.

tmmix. Advertisement. MultiLingual. April/May 2009: 38.

Wordfast LLC. "Wordfast." 29 May 2009.
<http://www.wordfast.net/index.php?whichpage=products&lang=engb>>.

Appendix C

Listing of Discussion Board Hierarchies

TranslatorsCafé.com (TranslatorsCafé.com 2009c)⁶⁶

Category	Forum	
General Discussion	Viewpoints, Questions, Comments, Advice Announcements Interpreting Working as a Freelancer Translators and Agencies From Language to Literature Hall of Fame and Shame All About Garbage	
Translation	Machine Translation Medical Translation TCTerms Features CAT and other Software for Translators Computer Work and Your Health Translator Resources and Web Links Recent Graduates: Ask Experts Techniques Ethics and Professionalism Jobs, Outsourcers and Payment	
Other Things We Can Do...	Localization Subtitling Voice-Over and Narration Audio Transcription	
Special	Off-Topic Chit-Chat Come Together Flea Market	
Language-Specific Forums	Albanian Arabic Bulgarian Chinese Czech Dutch English French German Indian Languages Italian	Japanese Other Languages Pidgins/Creoles Polish Portuguese Scandinavian Russian South Slavic Languages Spanish Turkish
TranslatorsCafé.com	Suggestions on Web Site Design Bug Reports and Troubleshooting Help! Testing Forum	

⁶⁶ Note that the discussion forums on TranslatorsCafé.com have undergone significant reorganization since the implementation of the methodology.

ProZ.com (ProZ.com 2009d)

Category	Sub-Category	Forum
Translation Art & Business	Art of Translation and Interpreting	Translator Resources Translation Theory and Practice Literature/Poetry Interpreting Subtitling Proofreading/Editing/Reviewing Translation Project/Vendor Management
	Business of Translation and Interpreting	Money Matters Getting Established Being Independent Business Issues Translation Agencies Translators Associations
	Language Studies	Linguistics Professional Development Teaching and Learning Languages Multilingual Families Ancient Languages Artificial Languages StudentZ Forum StudentZ Pub Linguistic Diversity
	Events and Awards	Language Industry Events & Announcements Honors/Citations 5 th ProZ.com Conference – Budapest 2007 2007 Regional Conference – Germany 2007 Regional Conference – Montreal 2007 Regional Conference – France 2008 Regional Conference - France TM Europe Conference – Warsaw Pact UK Conference 2009 Regional Conference – Germany 2009 Brazil Regional Conference 6 th ProZ.com Conference – Ohrid 2009 2009 Regional Conference – Poland 2009 Regional Conference – Austria 2009 Regional Conference – Argentina* 2009 Regional Conference – France* 2009 Regional Conference – Spain*
	Off-Topic/Other	Off Topic Lighter Side of Trans/Interp Translators4kids

Category	Sub-Category	Forum		
Technical Forums	Technical	General Technical Issues Localization Smart Shoppers Right-to-Left Language Technical Forum DTP – Desktop Publishing		
	CAT Tools Support	CAT Tools Technical Help SDL Trados Support Déjà Vu Support Wordfast Support Transit Support Alchemy Support NomaDesk Support Across Support OmegaT Support Metatexis Support MemoQ Support		
	Computing	Windows Operating Systems Apple/Mac Operating Systems Linux OS/Free Software Hardware		
	Software and Internet	Office Applications Software Applications Internet for Translators Safe Computing Speech Recognition		
Non-English Forums		Albanian Arabic Armenian Azerbaijani* Bengali Bosnian Bulgarian Catalan Chinese Croatian Czech Danish Dutch Estonian Finnish French Gaelic	Georgian German Greek Gujarati* Hebrew Hindi Hungarian Indonesian Italian Japanese Kazakh Kurdish Latvian Lithuanian Macedonian Malay Norwegian	Persian/Farsi Polish Portuguese Romanian Russian Serbian Slovak Slovenian Spanish Swedish Tamil Turkish Ukrainian Urdu Vietnamese
Country-Specific Forums		Translation in Canada Translation in Spain Translation in Uruguay		

Category	Sub-Category	Forum
Forums About ProZ.com	Site Features	KudoZ ProZ.com Job Systems Site Forums ProZ.com Directory ProZ.com Hosting Powwow Organizers ProZ.com: Translator Coop* ProZ.com Technical Support* Poll Discussion Translation Article Knowledgebase ProZ.com Membership ProZ.com Translation Contests
	Help Us To Improve	ProZ.com Suggestions ProZ.com Bugs
	Corporate Forums	Business Matters Operations/Tools Sales/Marketing Working with ProZ.com

The asterisk indicates that the forum was not present at the time of thread selection.

Appendix D

WordSmith Tools Word List for Keyword Selection

Graphical Word	Frequency	Lemmas	Exclusion Reason
TOOL	108	tool [31] tools [77]	general
TRADOS	93		brand name
CAT	50		
FILE	50	file [37] files [13]	general
WORD	47		general/brand name
TM	35	tm [31] tms [4]	
DOCUMENT	34	document [28] documents [6]	general
WORDFAST	29		brand name
SOFTWARE	28		general
VERSION	21	version [16] versions [5]	general
FORMAT	20	format [13] formatting [7] formatted [2] unformatted [1]	general
MACRO	18	macro [8] macros [10]	general
MEMORY	18	memory [16] memories [2]	
SEGMENT	14	segment [8] segments [6]	
WORKBENCH	14		brand name
TEMPLATE	12	template [9] templates [3]	general
CATALYST	10		brand name
COMPUTER	10		general
DEVELOPER	10	developer [1] developers [9]	general
MACHINE	10		general
PDF	9		general
RTF	9		general
SERVER	9		general
TAGEDITOR	9		brand name
COREL	8		brand name
INDESIGN	8		brand name
OFFICE	8		general/brand name
SDLX	7		brand name
TMX	7		

The rows highlighted in gray indicate the keywords chosen.

Appendix E

Forum Search Results

ProZ.com: Forum Selection					
Forum Name	Number of Keywords	Eliminated <i>A Priori</i>	Contained < 3 Keywords	Eliminated After Further Investigation	Included for Thread Analysis
2007 Regional Conference - France		x			
2007 Regional Conference - Germany		x			
2007 Regional Conference - Montreal		x			
2008 Regional Conference - France		x			
2009 Brazil Regional Conference		x			
2009 Regional Conference - Austria		x			
2009 Regional Conference - Germany		x			
2009 Regional Conference - Poland		x			
5th ProZ.com Conference - Budapest 2007		x			
6th ProZ.com Conference - Ohrid 2009		x			
Across Support		x			
Albanian		x			
Alchemy Support		x			
Ancient Languages		x			
Apple/Mac Operating Systems	4				x
Arabic		x			
Armenian		x			
Artificial Languages		x			
Being Independent	4				x
Bengali		x			
Bosnian		x			
Bulgarian		x			
Business Issues	4				x
Business Matters		x			
CAT Tools Technical Help	4				x
Catalan		x			
Chinese		x			
Croatian		x			
Czech		x			
Danish		x			
Déjà Vu Support		x			
Dutch		x			
Estonian		x			
Finnish		x			
French		x			
Gaelic		x			
General Technical Issues	4				x
Georgian		x			
German		x			
Getting Established	4				x

Forum Name	Number of Keywords	Eliminated <i>A Priori</i>	Contained < 3 Keywords	Eliminated After Further Investigation	Included for Thread Analysis
Greek		x			
Hardware	3				x
Hebrew		x			
Hindi		x			
Honors / Citations		x			
Hungarian		x			
Indonesian		x			
Internet for Translators	4				x
Interpreting	3				x
Italian		x			
Japanese		x			
Kazakh		x			
KudoZ		x			
Kurdish		x			
Language Industry Events & Announcements		x			
Latvian		x			
Lighter Side of Trans/Interp		x			
Linguistic Diversity	2			x	
Linguistics	4				x
Linux OS / Free Software	4				x
Literature / Poetry	3				x
Lithuanian		x			
Localization	4				x
Macedonian		x			
Malay		x			
MemoQ Support		x			
Metatexis Support		x			
Money Matters	4				x
Multilingual Families		x			
NomaDesk Support		x			
Norwegian		x			
Off Topic		x			
Office Applications	4				x
OmegaT Support		x			
Operations/Tools		x			
Persian/Farsi		x			
Polish		x			
Poll Discussion		x			
Portuguese		x			
Powwow Organizers		x			
Professional Development	3				x
Proofreading / Editing / Reviewing	4				x
ProZ.com Bugs		x			
ProZ.com Directory		x			
ProZ.com Hosting		x			

Forum Name	Number of Keywords	Eliminated <i>A Priori</i>	Contained < 3 Keywords	Eliminated After Further Investigation	Included for Thread Analysis
ProZ.com Job Systems		x			
ProZ.com Membership		x			
ProZ.com Suggestions		x			
ProZ.com Translation Contests		x			
Right-to-Left Language Technical Forum	0		x		
Romanian		x			
Russian		x			
Safe Computing	4				x
Sales/Marketing		x			
SDL Trados Support		x			
Serbian		x			
Site Forums		x			
Slovak		x			
Slovenian		x			
Smart Shoppers	4				x
Software Applications	4				x
Spanish		x			
Speech Recognition	3				x
StudentZ Forum	2			x	
StudentZ Pub		x			
Subtitling	3				x
Swedish		x			
Tamil		x			
Teaching and Learning Languages	2			x	
TM Europe Conference - Warsaw Pact		x			
Transit Support		x			
Translation Agencies	0		x		
Translation Article Knowledgebase		x			
Translation in Canada	3				x
Translation in Spain		x			
Translation in Uruguay		x			
Translation Project / Vendor Management	3				x
Translation Theory and Practice	4				x
Translator Resources	4				x
Translators Associations	3				x
Translators4kids		x			
Turkish		x			
UK Conference		x			
Ukrainian		x			
Urdu		x			
Vietnamese		x			
Windows Operating Systems	3				x
Wordfast Support		x			
Working with ProZ.com		x			
Total		96	2	3	28

TranslatorsCafé.com: Forum Selection					
Forum Name	Number of Keywords	Eliminated <i>A Priori</i>	Contained < 3 Keywords	Eliminated <i>A Posteriori</i>	Included for Thread Analysis
Albanian		x			
All About Garbage	5				x
Announcements		x			
Arabic		x			
Audio Transcription		x			
Bug Reports and Troubleshooting		x			
Bulgarian		x			
CAT and other Software for Translators	5				x
Chinese		x			
Come Together		x			
Computer Work and Your Health	4				x
Czech		x			
Dutch		x			
English		x			
Ethics and Professionalism	4				x
Flea Market		x			
French		x			
From Language to Literature	4				x
German		x			
Hall of Fame and Shame		x			
Help!		x			
Indian Languages		x			
Interpreting	3				x
Italian		x			
Japanese		x			
Jobs, Outsourcers and Payment	5				x
Localization	5				x
Machine Translation	4				x
Medical Translation	4				x
Off-Topic Chit-Chat		x			
Other Languages		x			
Pidgins/Creoles		x			
Polish		x			
Portuguese		x			
Recent Graduates: Ask Experts	4				x
Russian		x			
Scandinavian		x			
South Slavic Languages		x			
Spanish		x			
Subtitling	2		x		
Suggestions on Web Site Design		x			
TCTerms Features		x			
Techniques	4				x
Testing Forum		x			
Translator Resources and Web Links	5				x
Translators and Agencies	4				x
Turkish		x			

Forum Name	Number of Keywords	Eliminated <i>A Priori</i>	Contained < 3 Keywords	Eliminated <i>A Posteriori</i>	Included for Thread Analysis
Viewpoints, Questions, Comments, Advice				x	
Voice-Over Narration		x			
Working as a Freelancer	5				x
Total		33	1	1	15

Appendix F

Corpus Administrative Information

Forum	Thread title	Th #	Posting title	P Post #	Date	Author	Country	Languages	Posting
Smart Shoppers	Macintosh, CAT tools and technical (medical) texts	P30	Run windows inside OSX	P30-05	08/29/08	kimadil	United States	EN-AR	[body of posting]
Smart Shoppers	Macintosh, CAT tools and technical (medical) texts	P30	I'm a Mac user in the midst of the same issue	P30-07	08/29/08	Amy Duncan	Brazil	PG-EN	[body of posting]
Translator Resources	Trados 6.5 Freelance for SDLX 2005 or 2006	P01	Trados 6.5 Freelance for SDLX 2005 or 2006	P01-01	01/22/07	Eno Damo	Canada	EN-AL	[body of posting]
Translator Resources	Trados 6.5 Freelance for SDLX 2005 or 2006	P01	have you considered an upgrade?	P01-02	01/22/07	Elvira Stolanov	Luxembourg	GE-RO	[body of posting]
Business Issues	Why do Trados word counts and Original software word counts differ?	P02	Why do Trados word counts and Original software word counts differ?	P02-01	02/05/07	InfinityLS	United States	EN-FR	[body of posting]
Business Issues	Why do Trados word counts and Original software word counts differ?	P02	Solves a mystery	P02-02	02/06/07	Daina Jauntirans	United States	GE-EN	[body of posting]
Business Issues	Why do Trados word counts and Original software word counts differ?	P02	AnyCount vs. Workbench	P02-03	02/06/07	Claudia Alvis	United States	EN-SP	[body of posting]
CAT Tools Technical Help	XLF Help - target field	P03	XLF Help - target field	P03-01	02/13/07	JMBianco	United Kingdom	EN-SP	[body of posting]
CAT Tools Technical Help	XLF Help - target field	P03	Use an XLIFF editor	P03-02	02/13/07	Rodolfo Raya	Uruguay	EN-SP	[body of posting]
CAT Tools Technical Help	XLF Help - target field	P03	TagEditor not the right tool	P03-03	02/13/07	Piotr Bienkowski	Poland	EN-PO	[body of posting]
CAT Tools Technical Help	XLF Help - target field	P03	Try this...	P03-04	02/13/07	Jabberwock	Poland	EN-PO	[body of posting]
CAT Tools Technical Help	XLF Help - target field	P03	MS Word find/replace	P03-05	02/15/07	Samuel Murray	South Africa	EN-AF	[body of posting]

Corpus Content-Related Information

187

Corpus Classification Categories

Topic-Specific Categories	Brand-Name Categories	Operating-System Categories
AI - alignment	A - Across	WI - Windows
An - analysis	ACk - Across CrossTank	MA - MAC OS
BR - brand requirements	AL - Alchemy (incl. Catalyst)	LI - Linux
BST - buying/selling/trading TMs	AN - Anaphraseus	O - Other (Unix, Solaris, etc.)
CC - computer capabilities (memory, RAM, etc.)	AT - AppleTrans	
CO - concordancing	CC - CatsCradle	
CT - compatibility with other tools	CM - Comparator	
FCS - files created by TM system	DN - Dragon Naturally Speaking	
FO - formatting	DW - DWG Translator	
FST - source and target file formats	DV - DejaVu	
FTM - file format of TM	F - Fortis	
I - interface (w/n Word, own, GUI)	FD - ForeignDesk	
KS - keyboard shortcuts	FR - Fine Reader	
L - localization	FS - file searching software	
LA - languages	FU - Fusion	
MK - marketing	H - Heartsome	
MT - machine translation	IB - IBM Translation Manager	
N - networking (using tools on a network)	IH - in-house TM system	
NT - working with no tools	IW - Idiom WorldServer	
NU - new releases or upgrades	LB - Lingobit	
O - ownership	LI - Lingotek	
OF - other forums (e.g., Yahoo, Proz)	LO - Logoport	
OS - operating system	M - Metatexis	
OSr - open source programs	MI - Apple iWorks	
P - effects on translation process	MO - Microsoft Office	
PI - price (of tool)	MQ - MemoQ	

Topic-Specific Categories	Brand-Name Categories	Operating-System Categories
PjM - project management	MT - MultiTerm	
PM - payment	MU - MultiTrans	
PO - purchasing/obtaining software	OF - Olifant	
QU - quality of translations	OL - Open Language Tools	
S - segmentation	OO - OpenOffice	
SC - spell checker	OT - OmegaT	
SE - speed of translation	P - Passolo	
SP - support	PTL - +Tools	
SRX - Segmentation Rules eXchange	PT - PlusToys	
TD - trial or demo versions	RC - RC WinTrans	
TM - terminology management	S - Similis	
TMX - TMX files	SD - SDL, SDLX	
ToM - type of matches	SI - Sisulizer	
TP - technical problems	SW - Swordfish	
TR - training (on tools)	T - Trados/SDL Trados	
TC - text characteristics and format	TE - TagEditor (Trados)	
UM - usefulness of matches	TN - Transit	
VL - vendor lock-in	TO - TranslationOffice 3000	
VR - voice recognition	TT - Tortoise Tagger	
WC - word count	TV - TMXValidator	
	TW - Translator's Workbench	
	TX - TTXPress	
	W - Wordfast	
	WA - WinAlign	
	XB - Xbench	
	XE - XML editor	

Appendix H

Postings Cited

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P04-02	Localization	has anybody tried the new tool similis	Similis is really easy to use	02/26/07	Lucy-Jane Michel	05/04/09	http://www.proz.com/forum/localization/63730-has-anybody-tried-the-new-tool-similis.html
P07-03	Speech Recognition	Dragon Naturally Speaking and SDLX	Dictating works, macros are a different story	05/16/07	Reed D. James	05/04/09	http://www.proz.com/forum/speech_recognition/73180-dragon-naturally-speaking-and-sdlx.html
P08-02	CAT Tools Technical Help	Converting files and TMs from Trados to WordFast and viceversa	Depends on source file format	06/06/07	Heinrich Pesch	05/04/09	http://www.proz.com/forum/cat_tools_technical_help/74802-converting_files_and_tms_from_trados_to_wordfast_and_viceversa.html
P08-03	CAT Tools Technical Help	Converting files and TMs from Trados to WordFast and viceversa	Don't let the agency dictate	06/07/07	David Turner	05/04/09	http://www.proz.com/forum/cat_tools_technical_help/74802-converting_files_and_tms_from_trados_to_wordfast_and_viceversa.html
P12-03	CAT Tools Technical Help	Should CAT-tools be more consistent?	lock-in effect	09/15/07	Istvan Lengyel	05/05/09	http://www.proz.com/forum/cat_tools_technical_help/83924-should_cat_tools_be_more_consistent.html
P12-14	CAT Tools Technical Help	Should CAT-tools be more consistent?	Wordfast's problem	09/18/07	David Turner	05/05/09	http://www.proz.com/forum/cat_tools_technical_help/83924-should_cat_tools_be_more_consistent.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P16-08	CAT Tools Technical Help	how can I turn exported TM in TMX format?	TMX not compatible	11/26/07	Samuel Murray	05/05/09	http://www.proz.com/forum/cat_tools_technical_help/90274-how_can_i_turn_exported_tm_in_tm_x_format.html
P17-01	Speech Recognition	Need help: Dragon 9 + Transit XV = Problems with non-English characters	Need help: Dragon 9 + Transit XV = Problems with non-English characters	12/11/07	Claudia Alvis	05/05/09	http://www.proz.com/forum/speech_recognition/91663-need_help%3A_dragon_9_plus_transit_xv_problems_with_non_english_characters.html
P18-01	Translator Resources	Should I buy a tool like Trados?	Should I buy a tool like Trados?	12/31/07	Sonja Marks	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados.html
P18-02	Translator Resources	Should I buy a tool like Trados?	It depends on your customer mix	12/31/07	Latin_Hellas	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados.html
P18-03	Translator Resources	Should I buy a tool like Trados?	Some advantages, big expense	12/31/07	Kevin Fulton	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados.html
P18-04	Translator Resources	Should I buy a tool like Trados?	The underlying problem	12/31/07	ViktoriaG	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados.html
P18-05	Translator Resources	Should I buy a tool like Trados?	a must	12/31/07	Noe Tessmann	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P18-07	Translator Resources	Should I buy a tool like Trados?	My experience	12/31/07	Lesley Clarke	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-08	Translator Resources	Should I buy a tool like Trados?	Support for more than five languages?	01/01/08	Jeff Whittaker	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-09	Translator Resources	Should I buy a tool like Trados?	In my humble opinion...	01/01/08	gdesai	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-11	Translator Resources	Should I buy a tool like Trados?	User friendly?	01/02/08	Noe Tessmann	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-12	Translator Resources	Should I buy a tool like Trados?	My two ha'p'orth	01/02/08	Tony M	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-14	Translator Resources	Should I buy a tool like Trados?	Yes, a point in favour of W/f	01/02/08	Tony M	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html
P18-15	Translator Resources	Should I buy a tool like Trados?	Interesting question	01/03/08	Giles Watson	05/06/09	http://www.proz.com/forum/translation_resources/93005-should_i_buy_a_tool_like_trados.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P18-17	Translator Resources	Should I buy a tool like Trados?	Horses for courses	01/05/08	Nancy Burgess	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados-page2.html
P18-19	Translator Resources	Should I buy a tool like Trados?	Is Trados freelance really any good for freelancers?	01/07/08	Virginia Navascues	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados-page2.html
P18-20	Translator Resources	Should I buy a tool like Trados?	CAT tools are a must	01/08/08	Eleni Makantani	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados-page2.html
P18-22	Translator Resources	Should I buy a tool like Trados?	Try MemoQ	01/11/08	Rob Grayson	05/06/09	http://www.proz.com/forum/translator_resources/93005-should_i_buy_a_tool_like_trados-page2.html
P19-03	CAT Tools Technical Help	Trados or Wordfast?	Trados, Wordfast or Metatexis?	01/31/08	Michael Short	05/06/09	http://www.proz.com/forum/cattools_technical_help/95833-trados_or_wordfast.html
P19-04	CAT Tools Technical Help	Trados or Wordfast?	Trados compatibility	01/31/08	Selcuk Akyuz	05/06/09	http://www.proz.com/forum/cattools_technical_help/95833-trados_or_wordfast.html
P20-02	Getting Established	Which one ? Trados / wordfast / déjà vu ?	Don't - at least for the time being	02/05/08	Oleg Rudavin	05/06/09	http://www.proz.com/forum/getting_established/96203-which_one_trados_wordfast_déjà_vu.html
P20-03	Getting Established	Which one ? Trados / wordfast / déjà vu ?	Terminology and phrases	02/05/08	Victor Dewsbery	05/06/09	http://www.proz.com/forum/getting_established/96203-which_one_trados_wordfast_déjà_vu.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P21	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?				05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified.html
P21-04	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	They just want you to use that damn Trados thingy	03/02/08	The Misha	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified.html
P21-13	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	Practically it means	03/03/08	Heinrich Pesch	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P21-14	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	Some answers	03/03/08	Samuel Murray	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P21-16	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	Understanding CAT Tool requirement...	03/03/08	Kristina Mijic	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P21-17	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	real skill or paper?	03/03/08	Marius V	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P21-21	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	You'll note ...	03/03/08	Lawyer-Linguist	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P21-23	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	More than one kind of skill	03/03/08	Samuel Murray	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P21-25	Translation Theory and Practice	Applicants MUST purchase Trados/be Trados certified?	Was it a TRANSLATION job?	03/03/08	Jose Henrique Lamensdorf	05/06/09	http://www.proz.com/forum/translation_theory_and_practice/98617-applicants_must_purchase_trados_be_trados_certified-page2.html
P26-07	Localization	Translation software for MACINTOSH platform	Use Swordfish on Mac OS X	06/02/08	Rodolfo Raya	05/06/09	http://www.proz.com/forum/localization/106694-translation_software_for_macintosh_platform.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P26-08	Localization	Translation software for MACINTOSH platform	Wordfast and Office 2008 for Mac	06/03/08	Fabiana Zardo	05/06/09	http://www.proz.com/forum/localization/106694-translation_software_for_macintosh_platform.html
P27-02	CAT Tools Technical Help	How to reverse a translation memory?	Workbench help file	06/23/08	Giles Watson	05/06/09	http://www.proz.com/forum/cat_tools_technical_help/108517-how_to_reverse_a_translation_memory.html
P28-01	Business Issues	Translate more for less!	Translate more for less!	07/17/08	casey	05/06/09	http://www.proz.com/forum/business_issues/110440-translate_more_for_less.html
P28-03	Business Issues	Translate more for less!	I know which one you mean	07/17/08	ViktoriaG	05/06/09	http://www.proz.com/forum/business_issues/110440-translate_more_for_less.html
P28-05	Business Issues	Translate more for less!	I received it too	07/18/08	Claudia Alvis	05/06/09	http://www.proz.com/forum/business_issues/110440-translate_more_for_less.html
P28-13	Business Issues	Translate more for less!	The must is a TMX-compliant CAT tool - not Trados	07/18/08	ViktoriaG	05/06/09	http://www.proz.com/forum/business_issues/110440-translate_more_for_less.html
P29-02	CAT Tools Technical Help	Trados / Déjà Vu / WordFast	Quite a subjective choice	07/25/08	Wolfgang Jorissen	05/06/09	http://www.proz.com/forum/cat_tools_technical_help/111122-trados_déjà_vu_wordfast.html
P29-06	CAT Tools Technical Help	Trados / Déjà Vu / WordFast	Testdrive them	07/27/08	Wolfgang Jorissen	05/06/09	http://www.proz.com/forum/cat_tools_technical_help/111122-trados_déjà_vu_wordfast.html
P30-04	Smart Shoppers	Macintosh, CAT tools and technical (medical) texts	OK, I see what you mean	08/29/08	Marie-Helene Hayles	05/06/09	http://www.proz.com/forum/smart_shoppers/113902-macintosh_cat_tools_and_technical_medical_texts.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P31-01	Translation Theory and Practice	Advanced Translation Techniques: 10,000 words per day	Advanced Translation Techniques: 10,000 words per day	09/11/08	Mike007	05/07/09	http://www.proz.com/forum/translation_theory_and_practice/114967-advanced_translation_techniques%3A_10000_words_per_day.html
P31-02	Translation Theory and Practice	Advanced Translation Techniques: 10,000 words per day	My list of favourites	09/12/08	Peter Linton	05/07/09	http://www.proz.com/forum/translation_theory_and_practice/114967-advanced_translation_techniques%3A_10000_words_per_day.html
P31-10	Translation Theory and Practice	Advanced Translation Techniques: 10,000 words per day	VRS drawbacks	09/17/08	Tom Thumb	05/07/09	http://www.proz.com/forum/translation_theory_and_practice/114967-advanced_translation_techniques%3A_10000_words_per_day-page2.html
P38	Translator Resources	Trados + CD dictionaries with netbook?				05/07/09	http://www.proz.com/forum/translator_resources/125042-trados_plus_cd_dictionaries_with_netbook.html
P39-01	Software Applications	Creating a proofreading project in Trados professional	Creating a proofreading project in Trados professional	01/30/09	Matt Bissette	05/07/09	http://www.proz.com/forum/software_applications/126377-creating_a_proofreading_project_in_trados_professional.html
P42-02	Business Issues	What is the next best thing to Trados?	SDL Trados 2007 SP2	07/19/08	Jerzy Czopik	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P42-04	Business Issues	What is the next best thing to Trados?	Good agencies provide the software at no charge	07/19/08	Nicole Schnell	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados.html
P42-07	Business Issues	What is the next best thing to Trados?	Metatexis	07/19/08	Claudia Alvis	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados.html
P42-11	Business Issues	What is the next best thing to Trados?	Jerzy	07/19/08	Natalie	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados.html
P42-19	Business Issues	What is the next best thing to Trados?	Formats	07/19/08	Claudia Alvis	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page2.html
P42-21	Business Issues	What is the next best thing to Trados?	Pros and cons	07/19/08	ViktoriaG	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page2.html
P42-24	Business Issues	What is the next best thing to Trados?	Okay, let's get constructive	07/19/08	ViktoriaG	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page2.html
P42-29	Business Issues	What is the next best thing to Trados?	Metatexis	07/20/08	legal english	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page3.html

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
P42-30	Business Issues	What is the next best thing to Trados?	Metatexis and Trados	07/20/08	Martin Wenzel	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page3.html
P42-31	Business Issues	What is the next best thing to Trados?	At the risk of nitpicking...	07/20/08	Kevin Lossner	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page3.html
P42-34	Business Issues	What is the next best thing to Trados?	Thanks for the witty post, Natalie!	07/20/08	ViktoriaG	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page3.html
P42-39	Business Issues	What is the next best thing to Trados?	Yup, tried it myself and it's awesome value for the price	07/20/08	ViktoriaG	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page4.html
P42-41	Business Issues	What is the next best thing to Trados?	Slight disagreement	07/21/08	Janet Rubin	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page4.html
P42-42	Business Issues	What is the next best thing to Trados?	Please no more "CAT vs. no CAT" thing...	07/21/08	Tomas Cano Binder	05/07/09	http://www.proz.com/forum/business_issues/110584-what_is_the_next_best_thing_to_trados-page4.html
TC01-01	CAT and other Software for Translators	Do I really need TRADOS?	Do I really need TRADOS?	01/16/07	Nadine Buijs-O'Brien	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC01-02	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	Alex Eames	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC01-03	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	Nadine Buijs-O'Brien	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC01-04	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	Peter Linton	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC01-06	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	Terry Waltz	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC01-08	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	Terry Waltz	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC01-09	CAT and other Software for Translators	Do I really need TRADOS?	RE: Do I really need TRADOS?	01/17/07	James Bond	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8483&start=1
TC02-02	CAT and other Software for Translators	how do I get TRADOS?	RE: how do I get TRADOS?	01/30/07	Terry Waltz	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8563&posts=6

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC02-03	CAT and other Software for Translators	how do I get TRADOS?	RE: how do I get TRADOS?	01/30/07	Denis Hay	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8563&posts=6
TC02-04	CAT and other Software for Translators	how do I get TRADOS?	RE: how do I get TRADOS?	01/30/07	Inesa Borisko	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8563&posts=6
TC05-01	CAT and other Software for Translators	The Approaching Trados-Similis War	The Approaching Trados-Similis War	02/27/07	Lucy-Jane Michel	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8738&posts=4
TC05-02	CAT and other Software for Translators	The Approaching Trados-Similis War	RE: The Approaching Trados-Similis War	02/27/07	Nikita Kobrin	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8738&posts=4
TC05-03	CAT and other Software for Translators	The Approaching Trados-Similis War	RE: The Approaching Trados-Similis War	02/27/07	Terry Waltz	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8738&posts=4
TC07	CAT and other Software for Translators	Software for Collaborative Project				04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8900&posts=4
TC07-02	CAT and other Software for Translators	Software for Collaborative Project	RE: Software for Collaborative Project	03/26/07	Nikita Kobrin	04/15/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8900&posts=4

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC08-01	CAT and other Software for Translators	Everything you need to know about TRADOS	Everything you need to know about TRADOS	03/30/07	James Padley	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8953&posts=3
TC11-02	CAT and other Software for Translators	English-Persian trados 2007 translation software	RE: English-Persian trados 2007 translation software	05/09/07	Denis Hay	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9188&posts=3
TC13-03	CAT and other Software for Translators	Have you used "Across v.3.5"?	RE: Have you used "Across v.3.5"?	12/17/06	Nicole Keller	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8289&start=1
TC13-05	CAT and other Software for Translators	Have you used "Across v.3.5"?	RE: Have you used "Across v.3.5"?	04/30/07	Richard Delaney	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8289&start=1
TC13-10	CAT and other Software for Translators	Have you used "Across v.3.5"?	RE: Have you used "Across v.3.5"?	05/11/07	Laila Balke	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8289&start=11&posts=16
TC13-11	CAT and other Software for Translators	Have you used "Across v.3.5"?	RE: Have you used "Across v.3.5"?	05/15/07	Sima Witzig	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8289&start=11&posts=16
TC14-02	CAT and other Software for Translators	what is the software to use to build a translation memory ?	RE: what is the software to use to build a translation memory ?	05/29/07	Denis Hay	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9331&posts=2

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC15-01	CAT and other Software for Translators	Free Webinar on SDL Trados	Free Webinar on SDL Trados	06/05/07	Massimo Ghislandi	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9376&posts=1
TC16-02	CAT and other Software for Translators	How much can I charge for a TRADOS translation?	RE: How much can I charge for a TRADOS translation?	06/09/07	Hoang Hoang	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9405&posts=4
TC17	CAT and other Software for Translators	Trados troubles with Hebrew				04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9570&posts=2
TC18-02	CAT and other Software for Translators	Problem with DNS and WordFast	RE: Problem with DNS and WordFast	07/18/07	Derek Thornton	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9642&posts=5
TC21-02	CAT and other Software for Translators	OmegaT and TTX files	RE: OmegaT and TTX files	08/16/07	Denis Hay	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=9844&start=1
TC22	CAT and other Software for Translators	Calculating the percentage of 100% and fuzzy matches in OmegaT				04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10009&posts=5
TC24	CAT and other Software for Translators	TAG EDITOR - Introducing specific characters with "Find&Replace"				04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10098&posts=10

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC24-04	CAT and other Software for Translators	TAG EDITOR - Introducing specific characters with "Find&Replace"	RE: TAG EDITOR - Introducing specific characters with "Find&Replace"	09/25/07	Denis Hay	04/16/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10098&posts=10
TC27	CAT and other Software for Translators	TRADOS WORLD TOUR!!!				04/18/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=5102&start=1
TC27-08	CAT and other Software for Translators	TRADOS WORLD TOUR!!!	RE: TRADOS WORLD TOUR!!!	03/03/07	Prabhat Ghosh	04/18/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=5102&start=11&posts=15
TC29-02	Ethics and Professionalism	agency sent unlicensed Trados with instructions to crack it	RE: agency sent unlicensed Trados with instructions to crack it	02/01/07	Kamen Nedev	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8575&start=1
TC29-14	Ethics and Professionalism	agency sent unlicensed Trados with instructions to crack it	RE: agency sent unlicensed Trados with instructions to crack it	02/07/07	Kamen Nedev	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8575&start=11
TC29-15	Ethics and Professionalism	agency sent unlicensed Trados with instructions to crack it	RE: agency sent unlicensed Trados with instructions to crack it	10/31/07	Jonathan Ellis	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8575&start=11
TC29-18	Ethics and Professionalism	agency sent unlicensed Trados with instructions to crack it	RE: agency sent unlicensed Trados with instructions to crack it	11/16/07	Elizabeth Lyons	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8575&start=21&posts=27

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC31-01	CAT and other Software for Translators	TagEditor error msgs.	TagEditor error msgs.	12/02/07	Richard Markley	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10554&pos=ts=1
TC34-02	CAT and other Software for Translators	How to Find & Replace in Translator's Workbench?	RE: How to Find & Replace in Translator's Workbench?	12/20/07	Denis Hay	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10664&pos=ts=3
TC36-02	CAT and other Software for Translators	software for translators	RE: software for translators	01/21/08	Chen Zhonggiang	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10840&pos=ts=6
TC36-03	CAT and other Software for Translators	software for translators	RE: software for translators	01/23/08	Denis Hay	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10840&pos=ts=6
TC37-03	CAT and other Software for Translators	SDL Trados 2007 - Service Pack 2 released	RE: SDL Trados 2007 - Service Pack 2 released	01/27/08	Yana Lebedeva	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10869&pos=ts=5
TC39-02	CAT and other Software for Translators	Building only one TM ?	RE: Building only one TM ?	02/22/08	Laurent J Krauland	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11043&pos=ts=3
TC39-03	CAT and other Software for Translators	Building only one TM ?	RE: Building only one TM ?	02/22/08	Claude Morales	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11043&pos=ts=3

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC41-05	CAT and other Software for Translators	Wordfast	RE: Wordfast	03/07/08	Laurent J Krauland	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11150&pos=7
TC41-06	CAT and other Software for Translators	Wordfast	RE: Wordfast	03/07/08	Denis Hay	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11150&pos=7
TC43-01	CAT and other Software for Translators	SDL Trados upgrades	SDL Trados upgrades	03/22/08	David Angel	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11272&pos=5
TC43-02	CAT and other Software for Translators	SDL Trados upgrades	RE: SDL Trados upgrades	03/22/08	Nikita Kobrin	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11272&pos=5
TC43-04	CAT and other Software for Translators	SDL Trados upgrades	RE: SDL Trados upgrades	03/23/08	David Angel	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11272&pos=5
TC43-05	CAT and other Software for Translators	SDL Trados upgrades	RE: SDL Trados upgrades	03/23/08	Laurent J Krauland	04/20/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11272&pos=5
TC45-01	CAT and other Software for Translators	Talks on MemoQ: a next-generation TM CAT tool	Talks on MemoQ: a next-generation TM CAT tool	03/09/08	Nikita Kobrin	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11170&star=1

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC45-02	CAT and other Software for Translators	Talks on MemoQ: a next-generation TM CAT tool	RE: Talks on MemoQ: the most promising TM CAT tool	03/09/08	Terry Waltz	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11170&star t=1
TC45-09	CAT and other Software for Translators	Talks on MemoQ: a next-generation TM CAT tool	RE: Talks on MemoQ: the most promising TM CAT tool	03/11/08	Istvan Lengyel	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11170&star t=1
TC45-14	CAT and other Software for Translators	Talks on MemoQ: a next-generation TM CAT tool	RE: Talks on MemoQ: the most promising TM CAT tool	04/02/08	Denis Hay	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11170&star t=11&posts=18
TC48	From Language to Literature	Should using CATs for translating fiction be banned				04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=1
TC48-01	From Language to Literature	Should using CATs for translating fiction be banned	RE: Should using CATs for translating fiction be banned	04/26/09	Leonardo La Malfa	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=1
TC48-02	From Language to Literature	Should using CATs for translating fiction be banned	RE: Should using CATs for translating fiction be banned	04/26/09	Nikita Kobrin	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=1
TC48-03	From Language to Literature	Should using CATs for translating fiction be banned	RE: Should using CATs for translating fiction be banned	04/26/08	Marisa Pavan	04/21/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=1

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC48-05	From Language to Literature	Should using CATs for translating fiction be banned	RE: Should using CATs for translating fiction be banned	04/28/08	Salmaan Kureemun	04/30/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=11&posts=20
TC48-08	From Language to Literature	Should using CATs for translating fiction be banned	RE: Should using CATs for translating fiction be banned	04/28/08	Jonathan Ellis	04/30/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11494&star t=11&posts=20
TC51	CAT and other Software for Translators	Most economical way to use Trados ALTERNATELY in two computers.				04/30/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=11805&pos ts=4
TC54-05	CAT and other Software for Translators	Trados and Microsoft Office 2007	RE: SDL Trados 2007 and Office 2007	07/12/08	Denis Hay	04/30/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=8554&start =1
TC57-02	CAT and other Software for Translators	Installing Trados 2007 SP2	RE: Installing Trados 2007 SP2	07/29/08	Denis Hay	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12080&pos ts=6
TC60-02	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs	07/28/08	Laurent J Krauland	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=1
TC60-03	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs	07/28/08	Jonathan Ellis	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=1

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC60-05	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/28/08	Nanna Mercer	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=11
TC60-06	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/28/08	Laurent J Krauland	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=11
TC60-11	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/28/08	Laurent J Krauland	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=11
TC60-18	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/29/08	Ryan Ginstrom	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=21
TC60-19	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/30/08	Elizabeth Lyons	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=21
TC60-21	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/31/08	Harry Bornemann	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=21
TC60-24	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	07/31/08	Harry Bornemann	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star t=31

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC60-27	Working as a Freelancer	on translators' income and CATs used as a pretext to downgrade rates	RE: on translators' income and CATs used as a pretext to downgrade rates	08/22/08	Jacek K	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12073&star_t=51&posts=57
TC64-01	CAT and other Software for Translators	Why do I need translation tools?	Why do I need translation tools?	10/18/08	istvan kovacs	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12614&posts=6
TC69-02	Working as a Freelancer	How should I calculate a price, when the agency sent me a TM?	RE: How should I calculate a price, when the agency sent me a TM?	11/03/08	Thor Kottelin	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12704&star_t=1&posts=6
TC70-01	CAT and other Software for Translators	Translating .txt files using Wordfast	Translating .txt files using Wordfast	10/31/08	Katharina Loos	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12687&posts=9
TC72-02	CAT and other Software for Translators	Anyone know anything about Heartsome?	RE: Anyone know anything about Heartsome?	11/11/08	Denis Hay	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12753&posts=6
TC74-02	CAT and other Software for Translators	Wordfast/Trados license price comparison	RE: TRADOS LICENCE	11/26/08	Denis Hay	05/03/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=12864&posts=3
TC83-02	CAT and other Software for Translators	Mac + CAT	RE: Mac + CAT	10/18/08	Tom in London	05/04/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10920&star_t=1

Personal Posting #	Forum	Thread Title	Posting Title	Posting Date	Author	Access Date	URL
TC83-04	CAT and other Software for Translators	Mac + CAT	Mac community not big enough to justify dedicated Mac CAT tool	10/19/08	Jeff Allen	05/04/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10920&star t=1
TC83-07	CAT and other Software for Translators	Mac + CAT	RE: Mac community not big enough to justify dedicated Mac CAT tool	10/20/08	Jeff Allen	05/04/09	http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=10920&star t=1

Appendix I

Summary of Recommendations for Stakeholder Groups

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
PURCHASING OR ACQUIRING A TM SYSTEM					
<i>Situational Considerations</i>					
<i>Text Characteristics and Format</i>	investigate how TM systems may be useful to translate non-repetitive texts as well as repetitive texts evaluate the effects of TM usage on their workflow and translation process consider expanding current level of TM usage to include non-repetitive texts	research this area of TM usage as it may increase the marketability of their tool and allow them to promote to new groups of users or types of users		investigate the use of TM systems for non-repetitive texts to determine the types of users who translate all texts with TM systems, the perceived and attained benefits, brand-specific relations, method-based relations (i.e. CSB or SB), reasons for the appearance of this type of usage, and the effects on adoption investigate the current role of hard-copy document usage in non-adoption	not only address the use of TM systems for repetitive texts in training programs, but also introduce students to other potential applications of TM systems

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Languages</i>	investigate the availability of the interface in their first (or more comfortable) language and the potential positive effects on learning curve	research language-specific issues in order to determine the areas (or languages) in which users' needs are not being met and the effect of language support on brand choice		examine the extent to which the language of the interface (and supporting documentation) affects the learning curve in order to help translator trainers better address the needs of their students and provide further guidance for potential users when investigating TM systems investigate other sources to indentify the extent to which the language-related issues are being discussed and to determine whether or not the systems currently available have adequately addressed the issue	encourage students to consider the effect that the language of the interface, supporting documentation, and training materials may have on users' learning curve

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
Technical Considerations					
<i>Computing Requirements</i>					
Processing Speed and Memory	be aware of the importance of meeting vendor-specified requirements	be aware of the importance of providing easily accessible, up-to-date, and detailed information about technical specifications			
Operating System	continue to vocalize their OS preferences in order to ensure that vendors are aware that cross-platform tools are needed and being used	recognize that cross-platform capabilities may increase sales and research the potential for expanding adoption to users of other OSs		further investigate the OSs currently being used in order to help identify the demand for Macintosh- and Linux-compatible TM systems and establish if users' current OS needs are being met with the cross-platform tools already on the market	ensure students are aware of the availability of cross-platform tools

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Open-source Software</i>	investigate the availability of open-source tools (especially TM systems or TEnTs), and seriously consider these tools when choosing a TM system			investigate uptake of open-source applications by translators in order to determine how and why these programs are currently being used (or not used)	ensure students are aware of the availability of open-source tools, and the potential benefits and drawbacks of their use
<i>Word-Processor or Text-Editor Integration</i>	investigate the need for additional software, and how integration of a word processor or text editor affects the ease of use, learnability, and user-friendliness of the display	remain in contact with their customers to determine attitudes towards the current state of integration realize that users who are content with their current integrated systems may not easily migrate to a proprietary text editor		investigate this area further in order to better determine current attitudes towards integration versus proprietary editors and identify concerns to allow for a smoother transition	ensure students are aware of both types of systems (i.e. those with integrated word processors or text editors versus those with proprietary text editors) and the potential benefits and drawbacks of each option

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>TM Creation, Storage, and Usage Methods</i>	be aware of these different approaches (especially the SB CSB approaches) as one method may be more effective for their translation work	strongly promote CSB systems and the specific benefits as potential users are most likely unaware of the CSB approach		continue to investigate the benefits and drawbacks of methods in order to increase awareness and to aid translators when choosing a TM system examine current use of pretranslation and translators' attitudes in order to determine its role and implications in the industry	ensure students are aware of various methods for TM creation, storage, and usage (especially the SB and CSB approaches), and the potential benefits and drawbacks of each method
<i>Brand-Specific Considerations</i>					
<i>Trial or Demo Versions</i>	investigate and use trial and demo versions as their use can provide the most effective method for determining whether or not a particular brand is the most appropriate for their situation	continue to provide and promote trial and demo versions in order to aid potential users in making a decision consider the type of trial or demo version offered			ensure students are aware of the availability of trial or demo versions, as they may be extremely beneficial when choosing a tool

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Price</i>	not be guided solely by price when choosing a TM system: potential tools should be investigated and the decision should be based on numerous factors (including features, ease of use, and technical support) appreciate that the choice to use a TM system is the choice to invest time (and usually money) in a product with expected returns, and these returns are not necessarily realized in the short term	note that price is a significant motivating factor for potential buyers, and as such, the benefits of TM usage (and ROI) should be adequately addressed in promotional materials in order to better guide new and potential users	be aware of the potentially high price of many TEnTs and, in situations where particular brands or proprietary formats are required, consider options for helping the translator make the initial investment		ensure students are aware of the range of tools (and their prices), and encourage students to consider all aspects of a TM system when choosing a particular brand

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Updates or New Releases</i>	be prepared for a delay in the release of patches and service packs investigate the possibility of upcoming new releases and perhaps readjust time of purchase accordingly investigate the vendor's pricing policy for new releases, as this may affect the decision of which tool to purchase realize that TM systems are not a one-time investment; they must be updated like any other software application or piece of technology	continue developing patches and service packs as quickly as possible, and ensure that their customers are aware of the timeline for development investigate the attitudes of their customers and potentially adjust the product price or new release pricing policy consider supporting older versions of TM systems for longer, enabling both clients/agencies and translators to receive the necessary technical support and access newer file formats			

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
Using a TM System <i>Training</i>	consider the investment of time and money lost while learning when purchasing a tool and realize that their comfort level with TM systems and computers in general will affect their individual learning curve consider discussion boards as a source of help, but realize that they cannot provide adequate training reflect on their preferred method of training in order to ensure that they will have access to training that suits their learning style and needs offer feedback to the providers of the training	investigate the training needs of current and new users in order to provide the necessary means for users to learn how to use their tool seek feedback about training formats and materials from their customers and address the areas for improvement identified		continue investigating effective methods for the teaching of translation technologies	continue investigating effective methods for the teaching of translation technologies and ensure that course materials contain about the issues surrounding TM usage (as well as the features and functioning) remain informed of current uses, developments, and issues related to TM systems in order to adequately prepare future translators for the profession remain in contact with all stakeholders and explore new methods of incorporating multiple attitudes and viewpoints into training programs

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>File Formats</i>	ensure their TM system can handle and produce the necessary formats they commonly work with, and reflect on how their choice of tool may affect future work (e.g. the jobs they may or may not be able to accept, the ease of working with various file types)	continue to develop file format capabilities by increasing the number of formats which can be handled by a tool, creating filters to properly convert from one format to another, and addressing issues of formatting (e.g. font, styles, document layout)		continue investigating the formats used by translators in order to aid vendors and developers in determining the needs of translators and current deficiencies in file format support	
<i>Input and Output Files</i>					
<i>Compatibility among TM Systems</i>	investigate compatibility options (e.g. import and export capabilities) research a number of tools (all aspects of these tools) in order to make the decision based on what suits their needs best	support open standards and compatibility, as this may increase customer satisfaction and may facilitate the process of switching from one TM system to another thereby encouraging the use of multiple tools		investigate cross-over among tools and current strategies for dealing with compatibility issues	ensure that training materials (i.e. textbooks) address the issue of compatibility and that students are introduced to the concept of compatibility in training courses

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Vendor Lock-In</i>	be aware of vendor lock-in and perhaps consider various tools that could produce the necessary format(s) to determine which ones they may find more pleasing to work with	support open standards and compatibility, as this may increase customer satisfaction and may facilitate the process of switching from one TM system to another thereby encouraging the use of multiple tools			
<i>Open Standard Formats</i>	support the use of open standards, an effective and reliable method of exchanging data	support open standards and develop products which can read and produce open standard formats		explore the use of open standards to determine when they are being used (or not being used) by vendors, TM users, and clients/agencies, and to determine how open standards can better meet the industry's needs	ensure that training materials (i.e. textbooks) address open standards and that students are introduced to the purpose and applications of open standards in training courses

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Matches Proposed by a TM System</i>	experiment with the size of their TMs in order to determine whether one large TM or smaller specific TMs better suit their preferences experiment with the fuzzy match value to find the most appropriate threshold based on their editing versus translating preferences consider the usefulness of others' TMs and the amount of editing needed when using a TM from another source		ensure the TMs sent to their translators are accurate, up-to-date, and of high quality in order for them to be of assistance to the translators	investigate the relationship between the size and the source of TM and the quality of matches	address the strategies for TM division (i.e. one large TM versus smaller more specific TMs) in training courses and ensure that students are aware of the consequences of each strategy

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Technical Issues</i>					
<i>Customizability of TM Systems</i>	consider the importance of customization before choosing a TM system, as availability and level of customization differ from tool to tool	continue offering customization, or investigate how their tool could be customized in order to better meet the needs of the users			

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Technical Support</i>	investigate vendors' reputations for technical support, the format of support provided (e.g. phone, e-mail, online), and the price of technical support before choosing a TM system remain in contact with customer service departments in order to ensure that current technical issues and solutions are made known to the vendor consider discussion forums as a form of support, but first research the problem before turning to their peers continue to support each other and share advice and tips with their fellow forum members	maintain accurate and current knowledgebases and FAQs to guide users when encountering technical problems investigate their own reputations for technical support in order to address the issues as identified by their customers			

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Procedural and Textual Effects of TM Usage</i>					
<i>Speed</i>	consider all aspects of their translation process and context when deciding to use a TM system and recognize that the use of TM systems (and subsequent realization of the benefits) may require a long-term investment				
<i>Quality</i>	recognize that a tool is just a tool and quality is not a result of TM use, but of the translator's abilities		ensure that the TMs sent to their translators are accurate, up-to-date, and of high quality in order for them to be of assistance to the translators		
<i>Modifications to the Translation Process</i>	consider the procedural benefits as well as drawbacks (i.e. database maintenance and alignment verification)			investigate the procedural benefits of TM usage and the contexts in which these benefits may be most apparent	

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
Client Relations					
<i>Brand or Format Requirements</i>	investigate methods for compatibility to allow for the use of their preferred brands when the requirements of the client or agency do not		communicate to their translators the reasons behind their tool choice and perhaps offer concessions when the requirement cannot be directly fulfilled (e.g. use of open standards, help in making initial investment in tool)	explore clients' and agencies' attitudes and perceptions of the use of TM systems in order to determine their motivations and to help bridge the gap between translators and clients/agencies	ensure that training materials (i.e. textbooks) address the issue of client brand or format requirements and that students are introduced to this issue in training courses

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Payment</i>	establish rates which reflect the time involved in translating when using a TM system (evaluating and editing matches, administration and maintenance activities) discuss the issue among themselves, recognize the impact of their choices and not “undercut” competitors, and work together to establish industry standards in order to effect the change that some TM users seem to want participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs	participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs	realize the time investment needed when dealing with exact and fuzzy matches and negotiate rates that reflect this participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs	explore the state of payment practices in the industry in order to establish current rates of pay for matches found in TMs investigate the amount of time involved in editing exact and fuzzy matches (versus translating the entire segment) participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs	ensure that training materials (i.e. textbooks) address the issue payment practices for matches found in a TM and that students are introduced to this issue in training courses participate in the development of best practice guidelines to assist translators and clients/agencies in negotiating rates for matches found in TMs ensure students are aware of current challenges and present both sides of the situation

	Users should...	Vendors and developers should...	Clients and agencies should...	Researchers should...	Trainers should...
<i>Ownership</i>				further address issues related to ownership to help determine the importance of the issue for both TM users and clients/agencies	
<i>Buying/Selling/Trading TMs</i>				determine when TMs are being bought/sold/traded and why translators may not necessarily be discussing the benefits or drawbacks of these exchanges	