Expanding the Possibilities of Translation Memory Systems

From the Translator's Wishlist to the Developer's Design

by

Pelagia Maria Lagoudaki

This thesis pertains to translation technology,
requirements engineering and translation studies.

Offered for the degree of

Doctor of Philosophy of the University of London

and the Diploma of Imperial College London

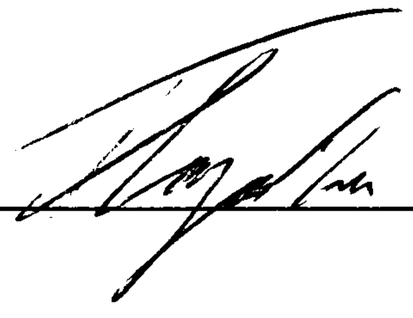
Department of Humanities
Imperial College London

IMPL. LONDON UNIV.

Author's Declaration

I confirm that:

- this thesis represents my own work;
- the contribution of any supervisors and others to the research and to the thesis was consistent with normal supervisory practice;
- external contributions to the research are acknowledged;
- parts of this thesis have been submitted and accepted for publication or published in advance of submission of the thesis.

Candidate:  Date: 1/12/2008

© 2008 Pelagia Maria Lagoudaki

Papers by the author which contributed to this thesis

1. "The Value of Machine Translation for the Professional Translator". In the proceedings of The Eighth Conference of the Association for Machine Translation in the Americas, Waikiki, Hawaii, 21-25 October 2008. Waikiki: AMTA.

2. “Translators evaluate TM systems - a survey”. MultiLingual Computing, No. 86, vol. 18, issue 2, March/April 2007, pp. 75-8.

3. “Challenges and Possibilities for Extracting Parallel Corpora from the Web – The Translator’s Dream Scenario”. In proceedings of the X International Symposium on Social Communication, Santiago de Cuba, 22-26 January 2007. Santiago de Cuba: Centro de Linguistica Aplicada, Ministerio de Ciencia, Tecnologia y Medio Ambiente, pp. 474-9.

4. “Translation Memories Survey 2006: Users’ perceptions around TM use”. In proceedings of the ASLIB International Conference Translating & the Computer 28, London, UK, 15-16 November 2006. London: ASLIB/IMI. [Also translated into Chinese by Su Myles]

5. “Technology in the service of the translation professional” (co-authored with Mark Shuttleworth). In proceedings of 1st Athens International Conference of Translation and Interpretation, Athens, Greece, 13-14 October 2006. Athens: Hellenic American Union.

Abstract

Translation Memory systems have become established as a key technology to address the challenges faced by the language services industry today. Despite flourishing research on how TM technology could reach the next level of technical sophistication, there has been a lack of systematic research on the usability aspects of TM systems, leaving many questions unanswered as to how translators' actual needs could be met more effectively by future TM solutions. The goal of this thesis has been to bridge the gap between current R&D efforts and translators' work reality by identifying and investigating the needs of modern translation professionals with reference to TM technology.

A survey was used to elicit information from TM users, with 874 translation professionals giving their views on what they liked, disliked and wished for in TM systems. The information was analysed using a system of coding and analytical elements of the grounded theory approach, as well as user profiling, in order to reveal the needs of various distinct groups of translation professionals, in relation also to particular TM use situations.

The identified needs relate to the functionality of TM systems, specifically to the TM translation editor, its facilities for resource building and management, project management and quality control, its search and translation assembly capabilities and its collaboration-enabling features. In addition, several other needs concern non-functional aspects of TM systems, namely the systems' usability. The thesis also explores those needs that cannot be derived immediately from articulated user requests or problems, but can be anticipated from a number of significant implicit clues. A number of such clues reveal a growing need for technologies that can access vast lexical and factual resources, or integrate such resources in a TM-enabled workflow. Such a subtle need is believed to hold the key to innovative TM systems.

Contents

Preface 12

INTRODUCTION 15

CHAPTER 1: TRANSLATION MEMORY SYSTEMS IN CONTEXT 23

1.1 TM SYSTEMS AND TRANSLATION TECHNOLOGY 23

1.2 TM SYSTEMS AND COMPUTER-ASSISTED TRANSLATION 27

1.3 THE ORIGINS AND CURRENT STATE OF TM SYSTEMS 32

1.3.1 Historical overview 32

1.3.2 Overview of current TM systems 37

1.3.3 TM system architecture 38

1.3.4. Key internal TM processes 40

1.3.4.1 Segmentation 40

1.3.4.2 Data storage 42

1.3.4.3 Match search and retrieval 42

1.3.4.4 Match assembly 44

1.3.4.5 Alignment 48

1.3.5 TM system generations 50

1.4 SOCIAL CONTEXT OF TM SYSTEMS 51

1.4.1 The language services industry 52

1.4.1.1 Significance of TM systems 54

1.4.2 The translation professional 55

1.4.2.1 Usefulness of TM systems 56

1.4.3 What drives TM system development 57

CHAPTER 2: PREPARING FOR USER NEEDS INVESTIGATION 59

2.1 LINE OF INVESTIGATION 60

2.2 NEEDS INVESTIGATION IN SOFTWARE ENGINEERING 63

2.2.1 Software development lifecycles 63

2.2.2 Requirements Engineering 65

2.3 DEFINITION OF TM USERS 73

2.4 METHODS OF DATA ACQUISITION 75

2.4.1 Sources of data 75

2.4.2 Available elicitation methods 77

2.4.2.1 Observation 77

2.4.2.2 Think Aloud Protocols 78

2.4.2.3 Scenarios (user stories) 78

2.4.2.4 Prototyping....79
2.4.2.5 Focus group discussion (brainstorming)....80
2.4.2.6 Interviewing....80
2.4.2.7 Survey research....82
2.4.2.7.1 Survey mode....84
2.5 CHALLENGES IN DATA ELICITATION....87

CHAPTER 3: ELICITATION OF INFORMATION THROUGH SURVEY....89

3.1 PILOT STUDY....90
3.1.1 Design & implementation....91
3.1.1.1 One-on-one interviews....92
3.1.1.2 Brainstorming....93
3.1.2 Study outcomes....95
3.1.2.1 Factors affecting user needs....95
3.1.2.2 Input for questionnaire design....97
3.1.3 Remarks on the research methods for the pilot study....99
3.2 SURVEY DESIGN....100
3.2.1 Questionnaire design....100
3.2.2 Technical features of the survey....105
3.3 SURVEY IMPLEMENTATION....105
3.3.1 Sampling....106
3.3.2 Survey promotion....108
3.4 SURVEY RESULTS....109
3.5 DATA FILTERING....110

CHAPTER 4: ANALYSIS OF QUALITATIVE DATA....112

4.1 GROUNDED THEORY....113
4.2 DATA ANALYSIS STRATEGY....114
4.2.1 Coding scheme....118
4.2.2 User profiling....121
4.2.3 Classification scheme....124
4.3 QDAS-AIDED ANALYSIS....126
4.4 CHALLENGES IN DATA ANALYSIS....129

CHAPTER 5: WHAT ARE THE ACTUAL NEEDS OF TM USERS? – PART I: FUNCTIONAL ASPECTS....132

5.1 FUNCTIONALITY....133
5.1.1 Translation editor....135
5.1.1.1 Translation process....138
5.1.1.2 Target text preview....141
5.1.1.3 Visibility of formatting tags and underlying code....141
5.1.1.4 Pre-translation....142

5.1.1.5 Preservation of text formatting and page layout 143
5.1.1.6 Productivity enhancement tools 144
5.1.2 Resource building *145*
5.1.2.1 Text alignment 146
5.1.2.2 Terminology collection 149
5.1.2.3 Terminology extraction 151
5.1.3 Resource management *153*
5.1.3.1 TM maintenance 154
5.1.3.2 Terminology maintenance 154
5.1.3.3 Organisation of resources & sharing 155
5.1.4 Search *156*
5.1.5 Translation assembly *159*
5.1.6 Project management *163*
5.1.6.1 Project setup 163
5.1.6.2 File format support 165
5.1.6.3 Version control 166
5.1.7 Quality control *167*
5.1.8 Collaboration *170*

CHAPTER 6: WHAT ARE THE ACTUAL NEEDS OF TM USERS? – PART II: NON-FUNCTIONAL ASPECTS 174

6.1 USABILITY 174
6.1.1 User interface *175*
6.1.2 Customisability *177*
6.1.3 Learnability *177*
6.1.4 Supportability *178*
6.2 EFFICIENCY 179
6.3 EXTENSIBILITY 180
6.4 PORTABILITY 181
6.5 RELIABILITY 181
6.6 ADDRESSING CONFLICTING NEEDS 182

CHAPTER 7: KNOWLEDGE RESOURCES AND INNOVATION IN TM SYSTEMS 186

7.1 ACCESS TO KNOWLEDGE RESOURCES 189
7.2 ACQUISITION OF KNOWLEDGE RESOURCES 191
7.2.1 Bilingual corpora *191*
7.2.1.1 Compiling parallel corpora 193
7.2.2 The web and its characteristics as a repository of texts *194*
7.2.3 Automatic extraction of parallel texts from the web *195*
7.2.4 Integrating a corpus extraction facility into a TM system *197*
7.3 THE CURRENT AND FUTURE ROLE OF TM SYSTEMS 200

CONCLUSION **202**

References 208

APPENDICES **219**

Appendix A: List of commercial Translation Memory systems 220

Appendix B: TM Survey 2006 web questionnaire (screenshots) 222

Appendix C: TM Survey 2006: Summary of quantitative results 231

Appendix D: Excerpts from TM Survey 2006 responses on TM systems' strengths and weaknesses 242

Appendix E: Excerpts from TM Survey 2006 responses on desired features 280

List of Figures

Figure 1: Quah's schema of Applied Translation Studies 24

Figure 2: Place of TM systems in Translation Technology 25

Figure 3: Basic processes in a simplified TM workflow 48

Figure 4: Typical software development lifecycle 63

Figure 5: Typical Requirements Engineering process 71

Figure 6: Strengths and weaknesses of survey modes 86

Figure 7: Needs classification in a software quality characteristics taxonomy in NVivo 128

List of Tables

TABLE 1: CONCEPTS RELATED TO TM SYSTEMS AND THEIR DEFINITION.....27

TABLE 2: PLACE OF TM SYSTEMS IN COMPUTER-ASSISTED TRANSLATION.....28

TABLE 3: TYPES OF TM MATCHES.....43

TABLE 4: IMPORTANT DEFINITIONS IN REQUIREMENTS ENGINEERING.....67

TABLE 5: SURVEYS ON THE USE OF TM SYSTEMS.....83

TABLE 6: ESTABLISHED PARAMETERS PREDICTIVE OF USER NEEDS VARIABILITY.....96

TABLE 7: EXAMPLES OF CODED UNITS OF INTEREST AND THEIR BASIS FOR CODING.....120

TABLE 8: ISO SOFTWARE QUALITY CHARACTERISTICS.....125

TABLE 9: EXAMPLES OF QUALITY SUB-CHARACTERISTICS DERIVED FROM IDENTIFIED NEEDS.....126

TABLE 10: COMMON ERRORS DETECTABLE BY A TM SYSTEM.....169

TABLE 11: POPULAR TRANSLATION RESEARCH METHODS BY ORDER OF PREFERENCE.....188

List of Abbreviations

ALPAC	Automatic Language Processing Advisory Committee
API	Application Programming Interface
ATC	Association of Translation Companies
CAT	Computer-Assisted (or Aided) Translation
CMS	Content Management System
CSB	Character-string-in-bitext
EAGLES	Expert Advisory Group on Language Engineering Standards
EBMT	Example-Based Machine Translation
EC	European Commission
eCoLoRe	eContent Localization Resources for Translator Training
GCM	Global Content Management
GUI	Graphical User Interface
HTML	HyperText Markup Language
IR	Information Retrieval
LISA	Localization Industry Standards Association
MT	Machine Translation
NLP	Natural Language Processing
OCR	Optical Character Recognition
QAS	Qualitative Analysis Software
RE	Requirements Engineering
SE	Software Engineering
SGML	Standard General Markup Language
SL	source language
SQL	Structured Query Language
SRX	Segmentation Rules eXchange
TB	Term Base
TBX	Term Base eXchange
TL	target language
TM	Translation Memory
TMX	Translation Memory eXchange
TT	Translation Technology
TS	Translation Studies
TU	Translation Unit
XLIFF	XML Localisation Interchange File Format
XML	eXtended Markup Language

Preface

This thesis marks the end of a long and eventful journey which began almost instinctively when I first entered the world of translation technology as a novice translator. I can clearly recall my first introduction to one of the earliest Translation Memory systems and the fascination I experienced, inspired not by the system's abilities at the time, but also by its potential as a powerful tool for translators. After that I worked with a great number and variety of TM systems (from in-house and off-the-shelf to experimental and open-source) and although my fascination with this technology remained strong, it was impossible not to be discouraged by the technology's limitations and shortcomings. Yet, the exact same limitations and shortcomings that at times drove me almost to the point of denouncing the technology altogether and abandoning the profession of translator – if such misery had to be accepted as a 'necessary evil' – also fired my imagination for better solutions and, eventually, prompted this research.

Improving TM systems is of obvious scientific interest, as they are currently considered as a crucial technology in addressing the challenges faced by the language services industry. At a more personal level, the motivation for this research also comes from a desire to know how such tools can be of greater help to modern translation professionals, and how technology can even empower them to make an unrestrained choice of the translation methods, strategies and tools they feel more comfortable with and which bring out the best of their skills.

This research has benefited from my continuous curiosity for TM systems and other computer aids, and has benefited from my experiences gained while working in different roles within the translation industry, amongst others, as a translator, translation company owner, project manager within a translation department, and purchaser of translation services. If it were not for these experiences, it would have been hard to understand each TM user's needs. And if it were not for the incessant interest in such fast-evolving technologies, it would have been very difficult to meet the most daunting challenge of all, which was that of keeping pace with a rapidly developing technology while conducting this research.

During this research, I was privileged to have met a number of remarkable people who helped me, each in their own way, to sail through rough seas and dangerous passages and reach my destination. To all these people, I am thankful.

In particular, I would like to express my gratitude to my supervisor, Mark Shuttleworth, whose expertise, understanding and patience added considerably to my experience as a doctoral student. I appreciate his vast knowledge and skill in translation studies and translation technology and his help in writing and presenting papers. His suggestions,

comments and editing assistance during the writing of this thesis greatly improved the final product.

A very special thanks goes out to Dr Vassilis Agouridas (University of Leeds), my mentor, who provided detailed criticism and analysis of my methods during the critical stages of this research. Were it not for his persistence, direction and lessons in self-control and self-discipline, I would still be lost in the sea of great ideas.

For his wisdom and kindness, I am eternally grateful to my second mentor, Prof. Ruslan Mitkov (University of Wolverhampton), whose ideas and comments made a great difference to this thesis. I am indebted to his generous contribution of time, resources and feedback, and I cannot thank him enough for inspiring and encouraging me to lift my work to new levels by offering my very best.

I also wish to express my warmest thanks to Dr Gabriela Saldanha (University of Birmingham) who reviewed parts of my thesis and offered valuable guidance and feedback. Her generous contribution of time, energy and insight enhanced my experience as a researcher.

Appreciation also goes to Prof. Harold Somers (University of Manchester) and Prof. Charles Drage (Imperial College) for their suggestions and valuable feedback during my transfer examination. Also, I wish to express my gratitude to Dr Ioannis Saridakis (Ionian University, Greece), my advisor, for his suggestions and provision of relevant literature.

I am grateful for ideas, comments and contributions from the following friends and colleagues: Magdalena Verheyen (for her insights and expert knowledge in Grounded Theory), Dr Stephan Kolitz (Harvard University), Dr Juan Lalaguna (Imperial College) and Christophe Declercq (Imperial College). It has been a great pleasure working with the department, staff, and students at Imperial College, and I wish to thank the departmental staff for all the instances in which their assistance helped me along the way. I am also indebted to Su Peneycad, Brett Harmony, David Dacam, Rocío Rodtjer and Noreen Griffin who proofread parts of this thesis and contributed helpful editorial comment. Thanks also go to Joseph Eliahoo from the Statistical Advisory Service at Imperial who provided me with statistical advice at times of critical need.

Thanks are also extended to my fellow researchers, Idalina Gomes, Ana Mialdea, Vincent Chang, Meng Ji and Panayiota Vatikioti for our philosophical debates, exchanges of knowledge and skills, and venting of frustration during my PhD, which helped enrich the experience. I am also thankful to my friends Dr Xin Xin, Claudia Pozo and Frantzeska Kolyda for their motivation and encouragement.

I recognize that this endeavour would not have been possible without the practical support throughout my research of my colleagues at Vodafone: Jaakko Anttonen, Karl Slootweg,

Alan Richardson, Anthony Davies, Phil Hornsey, Ray Yee and Gabriele Dangel. I am grateful for their understanding and for allowing me to fit work around my PhD studies.

No doubt, my greatest appreciation goes to the 873 people who participated in the survey and the eighteen people who took part in the pilot study, without whom this research would not have been possible, especially those anonymous translators who responded with humour and whose answers put a smile on my face in gloomy times during data analysis.

Finally, I would like to thank Dr Minako O'Hagan from Dublin City University and Eric Atwell from the University of Leeds for taking time out from their busy schedules to serve as my external examiners.

Above all, I would like to thank my family: my parents Tassoula and Yiorgos and brothers Mimis and Costis, for their love and untiring encouragement throughout this journey.

Conventions

- All trademarks and trade names of TM or other software, registered or not, mentioned in this thesis belong to their respective owners. The trademark symbol ™ or the registration symbol ® are used only at the first occurrence of a trademark.
- The use of either feminine or masculine pronouns (e.g. he/she, his/her) is consistent throughout every chapter and alternates chapter by chapter.
- In Chapters 5 and 6, wherever phrases or sentences appear within quotation marks and in italics, and are not attributed to any bibliographical source, they represent the thoughts, opinions and ideas of the anonymous respondents of TM survey 2006. Most of the verbatim responses cited contained language mistakes, the most serious of which were corrected by the author, so that the repeated use of (Sic.) in every quote was avoided.

Introduction

Over the last twenty years, the translation profession has felt a growing pressure to satisfy an ever-increasing demand for translations. This demand, stimulated in part by the new media of communication and the promotion of the value of multilingualism in modern societies, has motivated a quest for solutions that would help translation professionals to respond to the new challenges and would enable them to deliver high quality translation services at greater speed and lower cost. In the early years of informatics (i.e. the 1980s and 1990s), the deployment of technology seemed one of the obvious choices for many language service providers and translation departments within organisations. Those early adopters of translator-support technology quickly demonstrated the benefits of using such systems and as a result triggered a massive uptake of the technology by language service providers who wanted to access similar benefits and, at the same time, enhance their competitive edge.

Translation Memory systems are the finest example of translator-support technology as they form a comprehensive solution that deals with most of the tasks involved in a translation process. Today, TM technology is largely established in the language services industry as the mainstream translator-support technology and, because of this, it attracts a growing research interest from both academia and the industry.

TM systems are important from a social, economic and scientific perspective. The social significance of TM systems stems mainly from their ability to address pertinent challenges facing the translation profession. From a social point of view, TM systems are essential tools for both the industry as a group and the translation professional as an individual,

given that they can provide assistance in the practice of translation. From an economic perspective, TM technology is of significant value since it offers an effective way to produce translations fast and at low cost, without a corresponding drop in quality. Both language service providers and their clients can achieve significant cost savings from content re-use and the streamlining of the translation process when using a TM system. Scientifically, TM systems are also important because they can serve as a testbed for the application and integration of knowledge deriving from different disciplines such as Translation Studies (TS), Natural Language Processing (NLP) and Software Engineering (SE). The study of TM systems and its findings can conversely contribute to the theory of these disciplines and advance epistemic knowledge.

In recognition of the importance of TM technology, research and development in TM systems has incessantly revolved around a core question: how to improve TM systems. This theoretical question has also sparked the present research. The question can often arise from existing problems with TM systems, such as the fact that many systems appear to be complex and full of features with no real usefulness for a translator [Hunt, 2003], the fact that more than one TM system is often required to deal with different translation projects and the fact that many translation professionals are buying TM systems but never using them because of the steep learning curve involved [Lagoudaki, 2006].[1] Evidently, the current translator-support technology leaves a lot to be desired. However, the necessity of improving TM systems is not limited to addressing existing problems. The question remains relevant even after solutions have been offered for critical problems, because TM technology is dynamic, in the sense that it evolves along with the changing requirements of the relevant stakeholders, which constantly renew the scope for improvement.

The motive for pursuing an answer to this question varies according to the interests of each stakeholder who is engaged in TM research. For example, developers of commercial TM systems wish to improve TM systems primarily in order to achieve business and financial gains from selling a high-quality product. Developers of internal TM systems, on the other hand, seek to improve their systems in order to increase the productivity of their in-house translation teams. Our motive for pursuing the improvement of TM systems lies in the aspiration to see better translator-support tools that empower translation professionals, regardless of their organisational environment and its advantages or constraints, tools with a greater scope of utility that will appeal to a wider section of the translation community, and that will help them overcome in a more flexible way some of the challenges facing the profession.

[1] TM Survey 2006 revealed that translation professionals may use between one and eight TM systems for their work (53% of TM users reported using more than one TM system). The same survey also reveals that 16% of TM non-users have already bought a TM system, but they have not learned how to use it because of the complexity of the software.

Research area and the author's approach

TM systems' improvement is a research area that is being explored continuously. However, it is an area that is almost exclusively approached from the engineer's point of view. The existing literature on TM systems development shows that, up to now, technical feasibility has dominated the way to TM systems improvement, whereas systems' usability and end-users' demands seem to have been of only subordinate interest.

A review of several proposals by TM researchers either for the design of a complete TM system or for a single internal TM process reveals a repeated pattern of argument on which every proposal is constructed. The pattern involves the presence of an engineering problem within a TM system, which the TM researcher is called upon to solve. The research usually starts by referring to the technical limitations of other systems, before proposing a new approach.[2] This is a perfectly legitimate way of advancing knowledge as long as there is clear evidence that the limitation impacts on the quality of the system as it is perceived by its users. Since TM systems are tools in the hands of professionals rather than being mere works of engineering, such evidence cannot be provided by any other means than the users themselves. Unfortunately, rarely does a piece of research embark from evidence provided by TM users concerning an actual limitation of an existing TM system, and even when such evidence is presented, it is mostly anecdotal.[3] Moreover, research is hardly ever based upon a formalised user need. Instead, relevant needs are discussed in an expeditious and casual manner, while user-related information appears to be based on assumptions.

The motivation that underlies most of the reviewed research is the technical improvement of the TM system and not how the TM system can best meet the needs of its users. The perceived notion of how to improve TM systems seems then to coincide with improving processes like segmentation, alignment and matching. Indeed, an improvement in these processes may enhance the overall quality of the system to some degree, but research that involves a human evaluation of the implementation, capable of establishing such a relationship, has been scarce so far.

The lack of user-fed evidence for problems encountered in real-life conditions is not the only weakness of available research on TM development. The excessive focus on technology issues, instead of human needs, exemplifies a paradigm of technology-orientated research devoid of its social context. It is often apparent in the literature that TM systems are treated like Machine Translation (MT) systems, in the sense that excessive attention is attached to

---

[2] For example, Gábor et al. [2004] start their paper as follows:
"Commercially available translation memories do not involve linguistic knowledge. [...] existing Translation Memory systems do not systematically treat and recognize morpho-syntactic or even syntactic similarities, if the translation units themselves are too different in terms of characters. This paper presents a proposed system addressing these particular problems...".

[3] For example, in MultiCorpora's White Paper [2002], although the authors mention that "Many TM users report that they rarely use the concordance search features because of the very slow search speed" in order to argue about the limitations of existing TM systems, their statement is not supported by any evidence whatsoever. Therefore, the truthfulness of the fact that many TM users rarely use the concordance search features can be contested, and so can the reason offered – i.e. very slow search speed.

the automatic processes of the system, as if the success of the system depends on these processes only. In those cases, the bigger picture seems to be lost. MT systems have different objectives than TM systems, hence the focus of research is naturally on NLP problems. But TM systems are translator-support tools operating in constant interaction with the user with the aim of facilitating her work. Therefore, the main focus is probably misplaced if it is not on how to advance this interaction and how to identify ways of supporting the translator in a more effective way.

A piece of research that does not start from real-world facts in order to propose technological directions but departs from engineering challenges which call for solutions is likely to be carrying several risks. Apart from raising the question of whether the proposed solutions are actually desirable, there is a real danger that technological advances become so remote from the translators' work practice that ultimately they will be perceived as irrelevant to the translation task as performed by human translators.

Overall, research on TM systems (in general, and not only on TM system development) has traditionally focused on the tools themselves and not on the use of the tools in a broader context. Thankfully, in the past few years a slow but steady increase of interest in user issues has been observed, mainly amongst Translation Studies researchers and translation practitioners. This trend has been favoured also by a shift in focus in the discussion agenda of leading conference events (such as the ASLIB international conference 'Translating and the Computer') which have been encouraging the dissemination of TM user-orientated research over the past few years, and by the increased interest shown by industry associations (such as LISA) in user aspects through their surveys (e.g. LISA TM Surveys 2002 & 2004). There are already quite a few works on TM use that reveal the user's perspective. These are mainly divided into works that treat practical aspects of TM use, those that conduct user evaluations and those that investigate TM user perceptions.

Not many books have been written on practical aspects of TM use, but the ones by Bowker [2002] and Esselink [2003, 2000] are good examples of enlightening work that provides an insight into the context of use of TM systems and many of the surrounding practical aspects. The main bulk of information is found, however, in translation practitioners' journals (e.g. ITI Bulletin), translation technology-related magazines (e.g. MultiLingual) and conference proceedings. Some of the frequently discussed aspects include how TM systems are used in particular work situations [Zetzsche, 2006; Carter-Sigglow, 2004; Iverson, 2004; Musale, 2004; Joy, 2002; O'Leary, 2000], and TM ownership issues [Valderrábanos, 2004]. Other work offers product evaluations by users of TM [Garcia and Stevenson, 2007; Levitt, 2004; Gow, 2004; Benis, 2005, 2003a, 2003b], whereas more recent user-orientated research has focused on the general perceptions of translation professionals around TM technology, by either investigating user attitudes [Dillon and

Fraser, 2007] or by extracting information on TM usage and the factors that affect TM use [Lagoudaki, 2006; LISA, 2004, 2002; Wheatley, 2003].

Some of the above user-focused research could provide a fertile – though limited – basis for identifying user needs with reference to TM systems (see also section 2.1). There are often cases when users describe problems encountered while using a TM system and propose solutions that they think would fit their circumstances. There are also cases, scattered throughout the literature, where users visualise their ideal TM tool and suggest ideas. However, these problems or ideas remain hints of needs still to be discovered. They never lead to the actual definition of user needs, because, until now, user-focused research has lacked the formalism that would allow a systematic investigation of user needs.

The gap between social (user-focused) and scientific (technology-focused) research on TM systems is probably the main reason why the needs of the users struggle to find their way into development proposals. This thesis is an attempt to bridge this gap by approaching the question of TM systems improvement from the social scientist's point of view and by offering a systematic investigation of user needs which will supply the evidence needed by TM system engineers prior to engineering future technological solutions.

User needs identification with reference to TM systems is an unexplored area so far, as no substantial user-focused research with an explicit user needs orientation is available in the public domain. Yet, it is a vital line of research that provides the nexus between the people who create the systems and the people who use them, with the aim of TM systems improvement.

Possible impediments to needs-orientated TM development research arise mainly from two facts that may have discouraged research engagement so far. Firstly, the interdisciplinary nature of the area requires a sound background in translation studies, social research and computer science (computational linguistics, software engineering), as well as experience in the use of TM systems, in order for the researcher to be able to shed light on the question from all angles. Secondly, the non-existence of prior work on the subject in the public domain, as well as the reluctance of companies that develop TM systems and possess such information to share their data with the rest of the research community,[4] make it difficult for someone to carry out research outside a commercial context (e.g. in a purely academic setting).

The approach to TM system improvement (through user needs identification) taken in this thesis is very much in tune with the *user-centred design* philosophy proposed by Norman [1986: 31-61], that is increasingly applied to the design of software [Battle, 2005]. According to this, the design should try to optimise the system functionality around how

[4] Naturally, developers of commercial TM systems refrain from disclosing any invaluable data such as gathered user requirements, firstly because such data is bound by intellectual property rights, but most importantly because this is what ultimately provides and safeguards their competitive advantage.

people can, want to, or need to work, rather than forcing the users to change how they work to accommodate the system or function. Interestingly, the fact that translators are expected to change their working style and fit around the way the system works has already been remarked upon in the user-focused TM literature [Joy, 2002], and this is an issue that should be addressed.

The chief assumption underlying our approach is to consider the TM system as a *socio-technical system*. According to Somerville [2004:21-2], socio-technical systems differ from *technical systems* in that they may include one or more technical systems, but crucially also include knowledge of how the system should be used to achieve some broader objective. This means that the systems include people (the operators) as an inherent part of them. The user of socio-technical systems, thus, plays a critical role in the success of the system, so the possibilities for successful applications can only be secured through knowledge of user needs.

The present research forms part of a global research movement that promotes a *user advocacy* for all existing consumer technologies (e.g. mobile phones, computer operating systems and software). The goal of this movement is to re-sensitise software engineers to the world of the "average user". It is a movement that works towards the development of needs-driven applications in close cooperation with the end-users and not technology-driven ones that are likely to suffer from poor usability. The "average user" is the key reference point in the system's development. Average users neither want to know nor care how a program works; their expectation is that technology will work "as advertised" and "as expected".

Looking at translation professionals, the significant majority of TM users can be considered as "average users" since IT proficiency is irrelevant to how good translators they are, hence it should not be expected as a *de facto* skill. But in view of current complex TM systems and the resistance noticed by many (see Quah, 2006:18 for examples) to adopting the use of TM systems, the answer to the problem might not be urging translators to be trained on TM systems – the dominant view supported by almost all university translator training programs and researchers in translator training (e.g. Dillon & Frasier, 2007) – but urging TM system developers to create software for "average users".

Aim of research

The overall aim of this thesis is to identify, analyse and consolidate the needs of translation professionals with regard to TM technology with a view to improving TM systems. It is our hope that these needs will reveal areas of improvement and point to ways of expanding the possibilities of the existing systems.

More specifically, the present work aims to describe the circumstances in which particular needs emerge and at establishing relationships between needs and specific user groups. Furthermore, it aims to discover which parts of the translation process call for greater automation and which TM system attributes are most important to translation professionals.

The main aspiration of this research is to bring technology closer to humans, by providing constructive input that will enhance software engineers' understanding of the operational context of the system. It aims at complementing the research conducted by TM system engineers towards the development of user-adequate and need-oriented applications in the field of translator-support technology.

Finally, it is hoped that this research will contribute to the body of literature on Translation Technology, advance the state of practice in TM systems engineering and play at least a small part in shaping the future of TM technology.

Thesis outline

The thesis is divided into seven chapters. Chapter 1 provides an overview of TM systems, by describing their technical and social nature, and the progress in their development up to this day. Both the systems and their users, i.e. translation professionals, will be situated in the modern landscape where they interact and aspects of their relationship will be explored. The factors that drive the development of TM systems will be identified and an assessment of the current role of translation professionals in the overall system development process will highlight the underrepresentation of their needs.

Chapter 2 accentuates the importance of user needs and investigates existing approaches to user needs identification from several disciplines. A number of methods and techniques from Requirements Engineering and social research will be combined to devise an effective methodology for systematic user needs identification in the context of TM technology. The chapter will show that an emergent, exploratory, qualitative approach that gathers data through a survey is the most suitable to address the research questions.

Chapter 3 looks at the preparation for user needs investigation consisting of a data collection process in two stages. Firstly, the results of a pilot study designed to raise preliminary issues in eliciting information from TM users will be used to contribute to the design of the subsequent survey. Secondly, the elaboration of the survey will be discussed and its results assessed.

Chapter 4 describes the strategy followed for the analysis of the collected qualitative data, which is based on Grounded Theory principles, user profiling and theoretical coding. It will present the elaboration of a conceptual framework used to identify needs in the data and to

reify theoretical positions. A classification scheme for organising user needs will also be presented and some of the challenges presented during data analysis will be discussed.

Chapter 5 concentrates on the identified needs in terms of the functionality of TM systems. It will examine user needs relating to a TM's translation editor, its facilities for resource building and management, project management, quality control, its search and translation assembly abilities and its collaboration-enabling features. Needs will be explored with reference to use situations and particular sets of users, and will be juxtaposed to the available literature and research in the same areas.

Chapter 6 looks into the needs that concern non-functional aspects of TM systems, namely the systems' usability, with special reference to user interface and the systems' customisability, learnability and supportability, as well as their efficiency, extensibility, portability and reliability. It will raise awareness of conflicting needs not only among different groups of users, but also among users and developers, and will provide suggestions for addressing such conflicts.

Chapter 7 explores those needs that cannot be derived immediately from articulated user requests or problems, but can be anticipated from a number of significant implicit clues. It will show that these subtle needs may hold the key to innovative TM systems. The chapter will also envisage the role TM systems may play in the future not only in terms of supporting the translation process, but also in the context of translator training.

The thesis concludes with a summary of achievements and conclusion of the endeavour of identifying the needs of translation professionals with reference to TM systems with the aim of improving translator-support technology. The conclusion will demonstrate that better understanding of user needs reveals new areas of technology expansion and will show the way to designing TM solutions that fit around a translator's work, by putting forward a philosophy of software development based on communication, feedback, simplicity and respect.

Chapter 1: Translation Memory systems in context

Any attempt to investigate the needs of translation professionals in reference to Translation Memory systems must start from a sound understanding of the technology in question. A sufficient level of background information will help define the technical context and nature of the system, and explain its purpose of use, its functionality, and those characteristics that make TM systems special, compared to other translation support tools. The aim of this chapter is to provide a fundamental description of the characteristics of the domain under investigation.

1.1 TM systems and Translation Technology

Translation Memory systems as a subject of study fall within the young field of Translation Technology (TT). This is an interdisciplinary field that pertains to Translation Studies, Software Engineering, Computational Linguistics and Natural Language Processing.

According to the Translation Technologies Observatory[5], the field covers the following thematic areas: development, management and exploitation techniques for corpora; translation memories; machine translation; software localisation; documentation

[5] The Translation Technologies Observatory is a thematic network in the field of Translation and Information Technologies financed by the Ministry of Education and Science of Spain. Website: http://www.uem.es/web/ott/ingles/default.html [last accessed on 04/04/2007].

management; creation and management of terminological databases; pre- and post-editing techniques; controlled languages; and standards used in the translation industry.

The first attempt to depict schematically the relationship between Translation Technology and Translation Studies was made by Quah [2006: 42], who revised Holmes' schema of Translation Studies [Holmes, 2000: 172-85] by adding 'Translation Technology' as a branch of 'Applied Translation Studies', as shown in Figure 1.

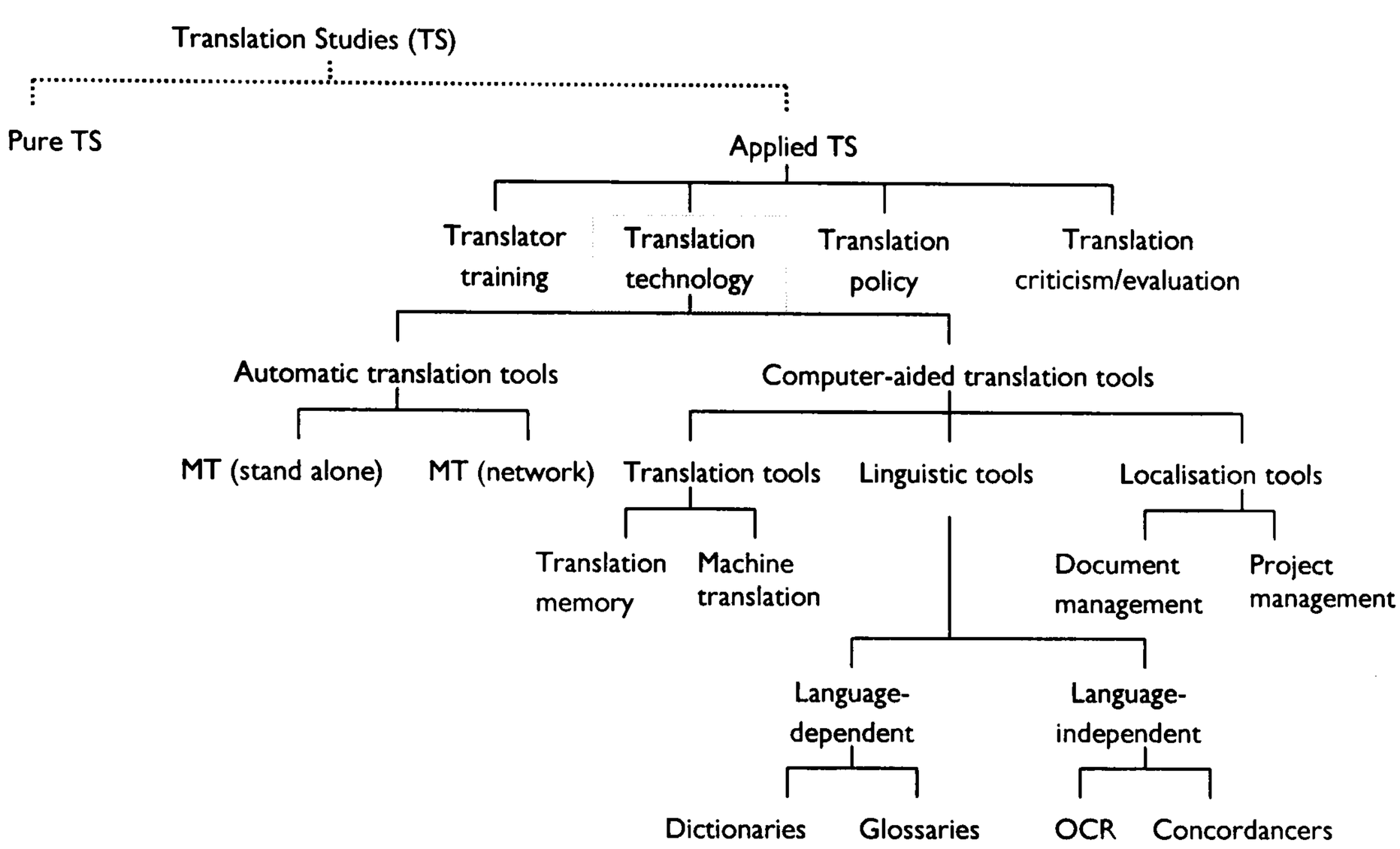


Figure 1: Quah's schema of Applied Translation Studies

According to Ulrych [2002: 200], Pure TS research aims for an improved understanding of languages, cultures and translation phenomena, whereas Applied TS uses the information obtained by Pure TS to address practical aspects of translation, such as translator training, translation tools, creation of translation policies and translation quality assessment. On this basis, Quah correctly identifies the place of Translation Technology under the Applied TS sub-field. However, in Holmes' original schema, there was an area called 'Translation aids', instead of 'Translation Technology', which included all the resources (such as dictionaries, glossaries and grammars) that translators normally use as translation aids. Naturally, no translation-supporting computer application was included in that schema because they were not identified as such yet (most of the current ones did not even exist). In Quah's updated schema, these resources are included as sub-branches of Translation Technology and are classified as language-dependent language tools [2006: 42].

One weakness of this classification is that it does not make any clear distinction between a *tool* and a *resource*. Dictionaries, glossaries and other reference material are widely

considered as a *resource* for translators [Bowker, 2002:7, Austermühl, 2001:8], in relation to which there are many *tools* available that people can use to create, manage and process these resources. Another weakness of Quah's classification of resources is that they cannot fall under the Translation Technology field since resources in general existed before any translation technology. They have certainly been enhanced (computerised) by translation technology and they have become more sophisticated, but it would be imprecise to say that translation technology has led to the appearance of resources. Therefore, it is deemed logical to favour the addition of a separate category – relevant to translation technology – under Applied TS, which would treat translation-relevant language resources as a stand-alone entity. This sub-field would study resources that serve as aids to translators and it would include both printed and digital resources such as dictionaries, glossaries, monolingual and bilingual corpora.

Quah's schema also presents a classification of translation systems based on the distinction originally made by Hutchins and Somers [1992: 147-8] and later refined by many, including Austermühl [2001: 8-11] and Hutchins [2004: 2]. In fact, virtually all researchers in TT who have offered proposals for categorising translation systems have based their typology on the Hutchins and Somers model [e.g. Sager, 1994: 290]. According to this model, fully automatic Machine Translation systems are distinct from translation support systems, in the sense that the former attempt to produce translations automatically without the intervention of a human, whereas the second category covers interactive tools that are designed to support the human translator during the translation activity [Hutchins, 2004: 2]. This distinction, based on the degrees of translation automation, is still relevant today and allows for a definition of a group of interactive translation support systems that includes Translation Memory systems, Terminology Management systems, concordance tools, localisation tools and a new type of translation workflow automation systems: Translation Management systems. The place of TM systems in relation to other Translation Technology systems is illustrated in Figure 2.

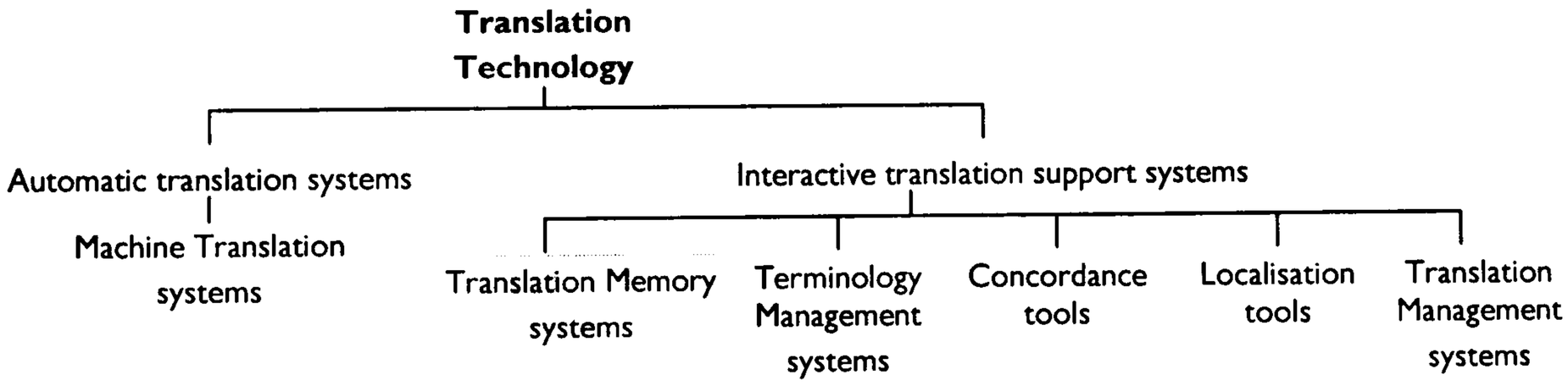


Figure 2: Place of TM systems in Translation Technology

It is important to note here that today it is quite common to see hybrid systems that are a combination of any – or all – of the above systems, depending on their intended purpose of use. It is also evident from the latest developments in those systems that translation systems, in general, have become multifunctional, which means that each system has integrated functions from different types of translation system or other systems outside translation technology. This may well be indicative of a possible future convergence of all translation systems into a single unified system, but at present we can still see in the market six separate types of translation systems (as shown in Figure 2) with distinctive characteristics. Broad definitions for each type of translation system, as well as definitions for concepts related to TM systems that will be used throughout this thesis, are provided in Table 1.

Concept	*Concept definition*
concordance tool (or concordancer)	A tool that provides a variety of information (e.g. context, occurrence frequency, collocations) on the use of a queried word or phrase in a series of texts residing in a database. It presupposes the existence of a database containing a (possibly bilingual) corpus. It is used by translation professionals, especially by translators and terminologists. It is also used by researchers in corpus linguistics as a text analysis tool to examine corpora, and by lexicographers.
Content Management System (CMS)	A set of tools that allow the creation, modification, organisation and publishing of digital content on a website. It is normally used by web authors, web editors and webmasters.
Example-Based Machine Translation (EBMT)	This term can refer to either the *technique* or the *system*. An EBMT technique is a approach originally used in Machine Translation systems, but now it is also being used in Translation Memory systems to produce 'machine translation matches'. It presupposes the existence of a repository containing paired 'examples' deriving from an aligned and segmented corpus. The technique tries to match the new source text against the repository of 'examples', to identify the corresponding translation segments (or sub-segments) and then to recombine these with the aim of offering a target text. An EBMT system is a Machine Translation system that employs the above technique.
Global Content Management (GCM) system	An enhanced CMS with the ability to store, manage and process multilingual content. It also offers translation workflow management functions. It is used by web authors, web editors, webmasters and translation professionals (e.g. translators, terminologists, project managers, localisers).
localisation tool	Software application used for the localisation of software, graphics and digital games. It is able to extract translatable text strings from the surrounding non-translatable code and it offers an editor for the processing of translation. It is also able to recompile the files by inserting the translated text back into the surrounding code. It includes a TM module, a terminology management module and various localisation workflow management facilities. It is normally used by localisers, software engineers, localisation managers and translators.
Translation Management system	Software application that incorporates a set of tools to support the translation process but gives greater emphasis on the management of the translation workflow within an organisation. They normally comprise a Translation Memory module, a Machine Translation module, a terminology management module and several project management facilities. It is normally used by translation companies or translation departments within an organisation.

Terminology Management System	Software application that includes a database (commonly called a *termbase* or *term bank*) where the user can store terminology, along with related information for each term, in a structured way. The system offers a set of tools to manage and process the terminology and its related information. Some systems are also able to extract terminology from texts (if sold separately, they are called 'terminology extraction' tools). It can be a standalone application or it can be seen as a component in a Translation Memory system. It is normally used by translation professionals, especially translators and terminologists, and by lexicographers.
Translation Memory (TM) system	Software application that includes a repository (most commonly in the form of a database) in which previous translations and their corresponding source text are stored in a structured and aligned way, so that any new text to translate is searched automatically and matched to the system's available resources in order for the system to be able to suggest a translation. The application normally comes with a set of standard tools (such as an alignment tool and various file format converters), as well as additional translation support tools (such as a terminology management tool, a concordance tool, etc.), thus offering an interactive translation support environment. It is normally used by translation professionals, especially by translators.
translation memory (or Translation Memory (TM) database)	A repository in which the user can store previously translated texts paired with their source text in a structured way. It can be a database within the system, or a repository ('virtual database') that includes associated texts that are located anywhere on one's hard disk, or an external database, located on a remote server. It is one of the two core components of a Translation Memory system (the other one is the search engine).
Machine Translation (MT) system	Software application that generates a translation for an input text automatically. It is used by anyone, and by translators as a translation support tool, when seeking translation suggestions for a source word, phrase or text.

Table 1: Concepts related to TM systems and their definition

1.2 TM systems and Computer-Assisted Translation

Since the focus of this thesis is the TM system from the user's perspective, it is essential to define the place of the system in what is particularly relevant to the translation professional and what has been called by translation researchers and practitioners Computer-Assisted Translation (CAT). Several definitions have been proposed for this concept, which have caused confusion around its meaning. Some define CAT in a narrow sense as a collection of translation tools (= *computer applications*, excluding computerised resources) [Bowker, 2002:7-8; Hatim and Munday, 2004: 336], and some even confuse it with a Translation Memory system [Benis, 2007a: 28-32; Samuelsson-Brown, 2004: 77-8; Valderrábanos, 2003:2].

For the purposes of this thesis, a broader definition of CAT is adopted to show the place of the TM system in the modern translator's comprehensive toolbox. In such a broad sense, CAT refers to translation in which *software applications* and *digital resources* are used by translation professionals to accelerate the translation process and guarantee or enhance the quality of their translation output. This collection encompasses a wide array of computer

applications either designed specifically for translation professionals or developed for general use, as well as language resources in digital form. Table 2 illustrates the place of TM systems in the collection of translators' aids, and reflects contemporary developments in CAT.

Digital resources	Computer applications	
	Translator-specific	General
Dictionaries & Thesauri Glossaries Termbases TM databases Monolingual corpora Bilingual corpora Encyclopedias Web	Translation Memory systems Terminology Management systems Translation Management systems Localisation tools (Concordance tools) (Machine Translation systems)	Word processors Webpage design/editing software Graphics processing software Spelling and grammar checkers Web search engines Desktop search software Desktop publishing software Data capture software (e.g. OCR) ....and many more.

Table 2: Place of TM systems in Computer-Assisted Translation

Many researchers have offered a similar typology of CAT aids in the past [e.g. Esselink, 2003: 79; Bowker, 2002: 7], which is now incomplete because, firstly, of the appearance of new kinds of tool that can be used by translation professionals today (such as desktop search software), and secondly because of the incorporation of long-existing tools, such as MT systems and concordance tools, into the translators' daily work after their potential for supporting the translation activity has been realised.

When examining the case of MT systems, it was long thought that MT is something that should be kept separate from translators, as it aspires to be able to produce the same quality of translation as humans do [Hutchins, 2001]. But things have changed lately. As Champollion [2003] and O'Hagan and Ashworth [2002: 43] notice, the use of MT is now considered a common practice in many organisations among translators who prefer to have a rough draft of a translation before they can produce a final translation by editing the first draft. This also turns out to be the case within the translation departments of the European institutions as suggested by information provided by the European Commission Directorate-General for Translation [2007:10-11].

As for the concordance tools that, according to Bowker and Barlow [2004:70-83], were once reserved for the academic community, they have been recently re-introduced to

translators by a number of TM system developers[6] who have integrated a concordance tool into their TM system, in order to compensate for the lack of context that the segmentation of texts is causing. The tool reveals contextual information for a suggested translation and this way it helps translators in making a more informed decision.

At the same time, the interest in concordance tools as translators' aids has been revitalised by the development of standalone bilingual concordancers geared towards translators, such as:

- WeBiText [Désilets, 2008];[7]
- RALI's TransSearch (now marketed by Terminotix) [Macklovitch et al., 2000c];[8]
- Multi-language Translation Example Browser [Goto et al., 2001/2003];
- WinDi® Translation Search Engine;[9]
- Linear B [Callison-Burch et al., 2004]; and
- Beetext Find.[10]

It is important to clarify here the difference between a bilingual concordance tool and a Translation Memory system, as the first one is often confused with the second, and is being referred to as a 'translation memory system' [Macklovitch, 2000; Macklovitch and Russell, 2000]. Besides any differences relating to the method of storing texts in their database – most TM systems store *segments* of text in a database, while concordance tools store *full texts* (bi-texts) – the key difference that draws a clear distinction between the two systems lies in their *search function*. While in a concordancer the user manually enters in the system the term, word or phrase he wants to find the translation for (*manual search*), in a TM system the search for each segment of the source text (associated or imported to the system) is done automatically (*automatic search*) [Bowker and Barlow, 2004: 71-7]. The idea behind these systems is that the user enters problematic expressions and the system returns from the bilingual database translation examples as results.

Apart from concordancers, another type of tool – this time one that is intended for general use – has recently found a place in the translators' aids collection, which can be used as a substitute for or in parallel with a Translation Memory system [Benis, 2007b: 26-29]. These are the Desktop Search tools,[11] and the fact that most of them are available for free may encourage their popularity among translators in the future. These tools index – usually in a fast and easy way – one's hard disk, which means, in the case of translators, that they can

---

[6] Examples of TM systems that include a concordance tool are: Wordfast 5.0 ('Contexts') and Déjà Vu X 7.0 ('Scan').
[7] Information about WeBiText can be found on http://www.webitext.com/ [last accessed on 25/09/2008].
[8] Information about TransSearch can be found on Terminotix' webiste: http://www.terminotix.com/eng/index.htm [last accessed on 30/05/2007].
[9] Information about WinDi Translation Search Engine can be found at: http://www.windi7.com/wse_info.php [last accessed on 30/05/2007].
[10] Information about Find can be found on Beetext's website: http://www.beetext.com/find.html [last accessed on 30/05/2007].
[11] Examples of such tools are: Google Desktop Search (http://desktop.google.com/) and Copernic Desktop Search (http://www.copernic.com/en/products/desktop-search/index.html) [last accessed on 30/05/2007]. Both tools are currently free.

have their old translations and source texts, their wealth of lexical resources (such as glossaries), any reference material (such as texts on different subjects), as well as any helpful e-mails from colleagues about the translation of particular terms, searched at the same time, from a single easy-to-use interface [ibid]. The advantages of this quick and easy information storage and retrieval process are clear over a TM system, and over a concordance tool, which they resemble more. However, Desktop Search tools lack the essential *cross-language matching function* that TM systems and concordancers have, i.e. the ability to match the search query of the user with its translation (if it exists) within the contents of their database.

It is interesting to notice here that nowadays translation professionals are in a more advantaged position, in terms of support, than ever before, thanks to the abundance of available aids spawned by the advent of computers. They are presented with a wide range of tools from which each professional can choose the tools that match better his needs and work style. But what makes Translation Memory systems more important than the rest of the available tools? On the surface, their popularity.

In the literature, it is widely acknowledged that TM technology has garnered unparalleled success among the translation community and has caused the greatest impact on translators' work, more than any other technology [Bowker and Barlow, 2004:71-5; Somers, 2003:31; Carl et al., 2002:19]. In reality, also, there is plenty of evidence for the success of TM systems. First of all, not only have they survived in a highly competitive translation software market for over 18 years, but their technology has also been evolving at a rapid pace thanks to the lively interest received by industry and academia. In addition, the increasing number and diversification of TM systems that are available on the market today undoubtedly reveals a rising demand for this type of technology. At the same time, surveys on the use of TM systems report a high penetration rate of TM technology (64% by eCoLoRe 2003, 91% by LISA TM Survey 2004, and 82% by TM Survey 2006)[12] as well as an increased TM usage rate [Wheatley, 2003, Lommel, 2004; Lagoudaki, 2006]. Finally, the popularity of TM systems over other translation tools can also be demonstrated by the large number of TM user groups that exist on the internet with many thousands of members.[13] Because of this evidence that makes TM systems stand out in the translators' toolbox, this thesis is specifically interested in the TM system rather than in any other system, since it appears to be the most popular translation support tool.

Furthermore, a more important reason behind our interest in TM systems is what explains the predominance of this kind of system. That is the *simplicity of the idea* underlying its operation (i.e. the ability of the system to offer translation suggestions from previously

[12] These percentages refer to the number of each survey's respondents who use TM systems.
[13] Searching Yahoo! Groups alone with the keywords 'translation memory' produces 31 TM user groups. As an indication, the user group of TRADOS has 4411 members and Wordfast group has 5025 members [search results accurate as on: 6/4/2007].

translated texts), and its *scalability*, meaning the flexibility of the system to incorporate functions from any other translation (or general) system that could render it a comprehensive and powerful translation support environment.

Many definitions have been offered for a Translation Memory system in the literature, each one reflecting the technology or the operation of the system that each person has in mind. Most frequently, a TM system is defined as an application whose technology enables the user to store translated phrases or sentences in a special database (called 'translation memory') for later reuse [Gotti et al. 2005:1; Benis: 2003; Esselink, 2000: 362]. However, since the appearance of TM systems (such as MultiTrans[14] and LogiTrans[15]) which allow for the storage of full texts in a database, this definition is no longer accurate. Moreover, some TM systems (such as STAR Transit[16]) do not even create a physical database, but instead make use of associated parallel texts, wherever these might be located in one's hard disk. In the light of these recent developments, the definition of TM systems needs to be more general to include all systems. For that reason, some researchers have defined TM systems in a broader sense as memory-based applications that allow users to recycle their old translations by comparing a source segment with an archive of aligned texts and matching it with the appropriate translation [Macklovitch and Russell, 2000:1; Macklovitch et al., 2000:1; Simard, 2003]. This generic definition makes no assumptions about the way in which the archive is queried, nor about the linguistic units that exist in the archive [Macklovitch and Russell, 2000:1]. However, it is so broad that it could include bilingual concordance tools.

This thesis adopts a narrower definition of a Translation Memory system (see also Table 1) that builds on the definition by Macklovitch and Russell and takes into account the automatic search function that distinguishes TM systems from concordancers. A TM system is, therefore, an application that links to a repository in which previous translations and their corresponding source text are stored in a structured and aligned way, so that any new text to be translated is searched for automatically and matched to the available resources associated with the system, in order for the system to be able to suggest a translation. The user of the system is then free to accept the suggested translation as it is, to edit it, or to reject it and translate the text himself. The application normally comes equipped with a set of standard tools (such as an alignment tool and various file format converters), as well as additional translation support tools (such as a terminology management tool and a concordance tool), thus offering an interactive translation support environment. This definition covers all systems that are based on the 'translation memory' concept (i.e. a system's ability to 'remember' previous translations), and that consist, among other components, of:

---

[14] Information about MultiTrans can be found at: http://www.multicorpora.com/index_e.html [last accessed on 30/05/2007].
[15] Information about LogiTrans can be found at: http://www.terminotix.com/eng/index.htm [last accessed on 30/05/2007].
[16] Information about STAR Transit can be found at: http://www.star-group.net/star-www/description/transit/star-group/eng/star.html [last accessed on 30/05/2007].

1. a *repository of resources* (= a translation memory), no matter whether those reside in a database (or databases) within the system, or anywhere in one's hard disk, or in an external database (located on a remote server), and
2. a *search engine* that automatically queries the resources and matches the source text with the appropriate translation (if it exists).

The architecture of a TM system is much more complex than that, therefore it will be discussed in more detail in section 1.3.3, but for the purpose of developing a definition for a TM system, which is the subject of this thesis, that is as precise as possible, it is important to have clarified the basic characteristics that all TM systems share.

The current definition is open enough to accommodate the existence of different approaches to TM technology and wide enough to encompass TM systems that have acquired such an advanced peripheral functionality that they are able to offer an integrated translation support environment. On this basis, our definition of a TM system is synonymous with what Hunt [2006] calls 'a Language Search Engine', what Zetzsche calls 'translation environment tools'[17] and what Reinke [1999] refers to as an 'integrated translation system'.

In fact, the current definition mirrors the current state of evolution reached by TM systems. The strategies to bring about better results have changed and multiplied through time, but the idea behind the systems has stayed the same since their inception. Interestingly, our definition is much closer to the original visualisation of a translators' aid system (or a 'translators' amanuensis'), as it was proposed by Martin Kay in 1980 [Kay, 1980/ 1997:3-23], than to many of the more recent definitions proposed by contemporary researchers.

1.3 The origins and current state of TM systems

1.3.1 Historical overview

Going back in time to trace the history of TM systems, it is hard to identify with certainty the origins of the idea for a translators' support tool, as different research groups have claimed originality. Even though most contemporary researchers attribute the original idea to Martin Kay, with his influential paper "The Proper Place of Men and Machines in Language Translation" published in 1980, Alan Melby [1995:225] suggests that the idea might have originated with his group at Brigham Young University in the 1970s. However,

[17] In his biweekly newsletter Tool Kit, no 71 (20/09/2006), Jost Zetzsche gives the following definition of a 'translation environment tool': "'Translation environment tool', [...]is a term that is geared toward tools with a translation memory component -- like Trados, Heartsome, Wordfast, across, MemoQ, Transit, etc. -- or a bitext component -- like Multitrans or Lingotek -- but also with tools for terminology management, analysis, code protection, project management, batch processing, spell checking, code page conversion, and many other features that most of these tools have."

the idea seems to predate any of their published works, as evidenced by the ALPAC report "Language and Machines", published in 1966, which, after discouraging any future research in MT, points to the development of computer-based aids for translators [Hutchins, 1998:3]. The ALPAC report was a decisive moment for the birth of TM technology, as it caused a shift in research and development from MT to more useful translation systems. More specifically, in this report, a system (developed by the European Coal and Steel Community) for 'automatic dictionary lookup with context' is described, which seems like an early incarnation of the 'translation memory' concept:

> The translator indicates, by underlining, the words with which he desires help. The entire sentence is then keypunched and fed into a computer. The computer goes through a search routine and prints out the sentence or sentences that most clearly match (in lexical items) the sentences in question. The translator then retrieves the desired items printed out with their context and in the order in which they occur in the source. [ALPAC, 1966:27]

Although this system was intended for terminological search, it included processes such as text alignment, as well as automatic matching and retrieval of terms in context, thus approaching the model of a modern TM system which still retains these essential processes.

Since then, many researchers elaborated on the idea of a terminological database, among whom was Fredrick Krollman, who in 1971 envisaged its further expansion into 'a linguistic data bank', which would include monolingual and multilingual dictionaries, thesauri, an index file, a corpus bank and translation archives [Hutchins, 1998: 5]. In the same year, Erhard Lippmann visualised a translators' support system where

> ...the system serves an extension of the capabilities of the user. [...] Rapid iteration toward the desired goal (i.e. a finished translation) can be achieved by switching back and forth as many times as required among human translation, direct dictionary lookup, editing, file management, and printing via typewriters or display screens. [Lippmann, 1971:10]

In fact, he went further into suggesting that 'access could also be made to computer resources outside the organisation, e.g. to remote terminological databases' [Hutchins, 1998: 6].

In 1979, Peter Arthern built further on the idea of a translation archive by proposing the development of multilingual text processing tools that would be based on what he called "translation by text-retrieval":

> The pre-requisite for implementing my proposal is that the text-processing system should have a large enough central memory store. If this is available, the proposal is simply that the organization in question should store all the texts it produces in the system's memory, together with their translations into however many languages are required. This information would have to be stored in such a way that any given portion of text in any of the languages involved can be located immediately... together with its translation... [Arthern, 1979: 94]

His vision of a tool that would search and retrieve 'any given portion of text' – in other words, one that is not limited to terms – stretched the functionality boundaries of the until then existing terminological tools and seems to be the earliest proposition for the

development of a tool that suggests translations for any source text from the texts that reside in its 'translation memory'. Arthern appears to be the first to make an implicit reference to a 'translation memory' as we know it today.

In 1980, Martin Kay published a paper in which he visualised in greater detail, with regards to functionality, a translator-support system called 'a translator's amanuensis'. His paper presented an advocacy for placing machines in the service of translators (as an opposition to the mainstream research efforts of that time, which were consumed for Machine Translation), and a proposition for the most complete version of a translator-support tool so far. He firstly proposed a 'special' multilingual text editor which would be the outcome of a 'flexible design' and that would 'make provisions for various kinds of extensions' [Kay, 1980/ 1997:14]. The editor's interface would be split into two windows: one that would contain the source text and another that would provide space for the translation. To this would be added a facility to automatically look up any word or phrase in a dictionary, and the ability to refer to previous decisions by the translator to ensure consistency in translation:

> ...the translator might start by issuing a command causing the system to display anything in the store that might be relevant to [the source text]. This will bring to his attention decisions he made before the actual translation started, statistically significant words and phrases, and a record of anything that had attracted attention when it occurred before. Before going on, he can examine past and future fragments of text that contain similar material. [Kay, 1980/ 1997:19]

Kay also envisaged an additional facility to his system: the automatic translation of text segments, which the translator could opt to let the machine do without intervention and then post-edit the result, or which could be done interactively, i.e. the system could ask the translator to resolve ambiguities [Kay, 1980/ 1997:20].

The most important contribution of Kay's visualisation of a translator-support system to the evolution track of TM systems is not the idea of the 'translation memory' *per se*, since that idea pre-existed (since Arthern). It is rather the unification of ideas around the design of a tool which for the first time attempts to address the needs of human translators. As Hutchins rightly notices, in retrospect,

> What Kay did [...] was to provide the impetus for the development of a workstation that incrementally incorporated aids that were really wanted by professional translators. [Hutchins, 1998: 9]

It is important to draw a parallel here between Kay's vision and the motivation behind this thesis. Just as Kay aspired to a way of bringing technology closer to translators' needs, this thesis is driven by the same aspiration. Our aim, though, is not a translator-oriented system design based on general observations about translators' work practices, as the one offered by Kay, but the systematic identification of the needs of contemporary translation professionals, that can form valuable input to a user-driven design of a TM system.

Interestingly, Kay's proposed design for his 'translator's amanuensis' is still considered a basic blueprint for TM systems [Somers, 2003b:32] and his prediction for the system's realisation is strikingly close to present reality:

> The translator's amanuensis will not run before it can walk. It will be called on only for that for which its masters have learned to trust it. It will not require constant infusions of new ad hoc devices that only expensive vendors can supply. It is a framework that will gracefully accommodate the future contributions that linguistics and computer science are able to make. One day it will be built because its very modesty assures its success. It is to be hoped that it will be built with taste by people who understand languages and computers well enough to know how little it is that they know. [Kay, 1980/ 1997:23]

In 1981, Alan Melby proposed his own model of a 'translator workstation', following the experience acquired when working on a multilingual MT system with extensive human interaction at Brigham Young University in the 1970s [Hutchins, 1998: 10]. His vision of a translator-support system was based on the idea of bringing together various ideas for supporting translators in an environment offering three levels of assistance [Hutchins, 1998:10]. At the first level, the translator would be able to use a text-processor to create his translation, supported by a terminology database and possibly a database of original and translated texts. At the second level, provided that the source text is in machine-readable form, the system provides also a concordance facility to allow the look up of peculiar words in their context. Finally, at the third level, the 'workstation' is integrated with a MT system that would be able to generate translations automatically, which translators would have the option to reject or incorporate into the target text unchanged or after editing [Melby, 1982:219]. Melby's proposed model coincides with Kay's earlier views, in that both stress the importance of human translators being in control of the system.

One of the earliest implementations (mid 1980s) of the previous ideas was the "repetitions processing" facility of the ALPS system. As Hutchins describes

> Translators could specify that a 'repetitions file' be created as the translation process proceeded. As each segment of text was translated, it was copied with its translation into the file. Segments could be multiple-word phrases or longer syntactic units. Once in the file, the translator could compare a new text against files of segments from already translated texts, extract those that matched and then create a file of similar segments for use on the text currently being translated. The translator did not have to retranslate repeated parts of texts. This facility was called "repetitions extraction" (abbreviated as 'repstraction'), and was clearly an early implementation of basic features of a 'translation memory'. [Hutchins, 1998:12]

During the same period, many more systems were developed, specifically designed for professional translators, which included some of the basic components of today's TM model, such as a text-processor, an automatic dictionary lookup facility and a concordance tool [Hutchins, 1998: 13]. However, a comprehensive translator-support system was still nowhere to be seen.

Things changed in the late 1980s and early 1990s, when advances in computer hardware and software met with important research outcomes in MT and corpus linguistics.

Programs moved from DOS to Windows, user interfaces became more standardised and the processing and storage power of PCs improved significantly. Advances like these, coupled with an increased availability of large electronic corpora of bilingual texts [Hutchins, 2003:510], generated new conditions for research, so that ideas like the one of 'bi-text' by Brian Harris [1988] could be realised. His idea essentially involved the existence of a bilingual database where the translator could store translations and their originals, segmented in translation units. Each translation unit of original text would be aligned to a unit that represented its translation. This way, the system would look for 'bi-text segments that will be similar enough to help him towards his objective' [Harris, 1998:9]. His 'bi-text' proposal is today seen as the predecessor of the bilingual concordancer, but it is also a refinement of the earlier ideas on the translation memory. It was, however, the "development of statistical means of text alignment" that made possible the "realisation of these and earlier embryonic proposals for bilingual databases of translations", as Hutchins rightly observes [1998:14].

In the early 1990s, while research and in-house systems continued to progress, the first commercial systems made their appearance. In 1991, the first commercial translator-dedicated system running on DOS was made available by STAR ('STAR Transit').[18] Soon after, in 1992, TRADOS released their commercial TM system running on DOS,[19] incorporating a 'translation memory' and alignment facilities, under the name "Translator's Workbench". According to Hutchins [1998:15], TRADOS appeared to be the first to use the term 'translation memory' for its product. In the same year, IBM introduced onto the market 'TranslationManager/2'.[20] A year later, ATRIL released 'Déjà Vu' for Windows 3.1,[21] and the following year TRADOS introduced 'Translator's Workbench' for Windows, while at same time STAR released the Windows version of their system 'STAR Transit'.[22] From 1994 onwards, the TM software market blossomed. New TM systems appeared year after year, most surviving up to the present day, while others enjoyed a promising start only to fall into oblivion a few years later.

A landmark development in the TM technology of the late 1990s was the remote access of a translation memory, enabled by the internet. Logoport Software was the first developer to launch a commercial Web-based Translation Memory system ('Logoport') that allowed Logoport's client companies to have online access to a central TM database that resided on

---

[18] This is according to combined information from 1) *Wired* magazine 8.05 (available at: http://www.wired.com/wired/archive/8.05/timeline_pr.html) and 2) the University Johannes Gutenberg's list of 'translation-relevant software (available at: http://www.fask.uni-mainz.de/ze/term/cat-pas.html).

[19] This is according to 1) an unknown author in the July-August 1994 issue of Language Industry Monitor (article available at: http://www.lim.nl/monitor/trados-2.html) who quotes Jochen Hummel (co-founder of Trados): "1992 was the year we released MultiTerm for Windows and our DOS Translator's Workbench package", and 2) a document provided by Trados (accessible through http://www.lai.com/tmcompet.htm) where it states that: "In 1992, TRADOS ships MultiTerm for Windows and Translator's Workbench 2 for DOS".

[20] According to the University of Leipzig's list of terminology management tools (available at http://www.uni-leipzig.de/~xlatio/software/soft-termiman.htm).

[21] According to ATRIL's website (http://www.atril.com/aboutus.asp?InFrame=true).

[22] According to the University Johannes Gutenberg's list of 'translation-relevant software (available at: http://www.fask.uni-mainz.de/ze/term/cat-pas.html).

their servers. Since then, the modes of accessing TM databases have evolved from purely standalone to distributed data either synchronised with a central database or as a remotely accessed central database [Yuste, 2004:9-10].

It is, finally, worth noticing in the evolution of the TM marketplace the considerable interest shown in developing open source TM software. Some of the efforts have led to the release of working products (such as Transolution, ForeignDesk, OOxlate), and some others have been less fruitful. Among the open source TM systems, the most successful ones, which have been under continuous development up to the present day, are Omega-T and the Open Language Tools [Lagoudaki, 2006].

1.3.2 Overview of current TM systems

These days, more than ever before, a great number and variety of TM systems are available to translation professionals, with new systems entering the market every few months, intensifying the competition in the TM systems arena. At present, a quick scan of the TM software market will reveal more than 30 available commercial TM systems, an indicative list of which is provided in Appendix A. According to the ICL TM Survey 2006, the most popular TM systems with freelance users of TM are: Déjà Vu, WordFast, SDL Trados, STAR Transit, MultiTrans and Omega-T [Lagoudaki, 2006]. TM systems, nowadays, come in a selection of types. A common typology promoted by most TM developers is based on the degree of scalability of the system, reflecting differences in the intended use of the system by specific user groups, based on their particular needs. In such a typology, TM systems come in the following editions:

1. corporate systems, which include:
    - systems for Language Service Providers (translation/localisation companies)
    - systems for organisations dealing with multilingual content
2. systems for freelancers

Besides the most obvious difference of price (e.g. WordFast is cheap and Omega-T free, while most others sell at full commercial prices), TM systems generally differ in terms of:

- the graphical user interface (GUI) design;
- the architecture (which is discussed further in section 1.3.3);
- the key internal TM processes (discussed further in section 1.3.4), in particular in terms of:
    - the source file formats they support
    - the granularity of segmentation

    - the indexing method
    - the structure of the resources repository
    - the match search and retrieval techniques
    - the integration of machine translation techniques;
- the ability to port to other applications, such as Content Management Systems, authoring systems and Machine Translation systems; and
- the level of customisation allowed to the user.

The differences in GUI relate to the text editing environment in which users are allowed to process their translation. Some TM systems work as add-ins (through macros) to Microsoft Word, while others provide their own translation processing environment, usually in a tabular way. TM systems that belong to the first category are: Wordfast, MultiTrans, Metatexis, Trados and Fusion. DéjàVu, Heartsome, MemoQ, STAR Transit, SDLX and across are some of the systems that belong to the second category [Shuttleworth and Lagoudaki, 2006]. Both these approaches to GUI design have respective advantages and the choice of one design over the other is usually a matter of the user's preference [Zetzsche, 2005:164-7].

In conclusion, modern TM systems vary mainly in design and technology employed. However, their development always departs from the very same 'translation memory' concept, and for that reason they are largely based on the same underlying principles.

1.3.3 TM system architecture

TM systems are no longer just databases. As the name 'system' suggests, they may consist of several components linked together within a multi-layered software architecture. A typical TM system has a search-oriented architecture. The architectural layer is usually occupied by a database management system (DBMS), which is supplemented by a search engine. Queries for information are usually performed using Structured Query Language (SQL). The database management system comprises all those algorithms that govern internal operations related to the creation, manipulation and maintenance of the database, such as *text segmentation*, *alignment of translation units* (TU) and *storage*. Typically, two databases are created: a TM database and a terminology database, but the creation of additional databases containing other resources, such as a lexicon, is also possible in some systems (e.g. Déjà Vu X).

TM systems may have a multi-tier architecture, in which information is maintained in a Data-Tier where it can be stored and retrieved from a database or file system. This Data-Tier may be replaced or placed behind another tier which contains a search engine and search engine index which is queried in-place of the database management system. The

search engine itself crawls the database management system in addition to other data sources such as file systems and consolidates the results when queried. Apart from the two core components and their modules, TM systems also commonly include a text-editing module, file filters, and workflow mechanisms.

Recent developments in TM technology which caused the emergence of hybrid systems combining TM with MT technologies (e.g. Déjà Vu X) have led to the expansion of the search engine's operations, typically limited to *match search and retrieval*, to include an additional operation (previously performed only by MT systems): *match assembly* (or construction). This operation poses several special requirements to the TM system architecture, as it necessitates special modules to be added, such as a match construction algorithm and possibly POS taggers and parsers and various auxiliary lexical resources.

The same happens when a TM system combines technologies from other systems, such as localisation tools or term extraction tools. Then, the TM system acquires modules from these systems, and its architecture becomes more complex.

Regarding the architecture model, TM systems can be either centralised ('desktop systems', i.e. all system components reside on a single centralised computer and all the processes run there) or distributed systems. In the latter case, we can come across two types of TM system architecture:

a) *client/server*: The server side hosts the DBMS and the search engine, with all their modules, as well as any filters and workflow mechanisms. The created databases, which constitute the TM repository, are also hosted on the server, and the system enables their sharing, as well as their remote management and processing. The server may be located on the TM developer's computers, in which case it is accessed by the users via the internet ('Web-based TM') (e.g. Logoport), or it may be located on a company's server network, e.g. LAN (e.g. across Language Server). The client side is an interface (web browser, in the case of web-based TM systems) usually hosted in browser-based forms or Java applets. It provides the text-processing components, as well as the necessary code to communicate with the shared resources on the server [Savourel, 2001:329].

b) *service-oriented*: In 2006, Lingotek pioneered a service-oriented architecture for their commercial TM system, the result being a TM solution offered as a *web service* rather than as a *product*. Unlike traditional client/server models, web services do not provide the user with a GUI; instead, they share business logic, data and processes through a programmatic interface (e.g. web browser) across the internet. Developers can then add the web service to a GUI (such as a web page or an executable program) to offer specific functionality to users.[23] For a TM user this practically means that he has access to a TM system through the internet, instead of owning one. The main advantages of this solution are that the TM

[23] Definition according to Webopedia (available at: http://www.webopedia.com/DidYouKnow/Computer_Science/2005/web_services.asp) [last accessed on 12/09/2008].

system can be accessed from any platform (e.g. Windows, Mac, Linux) and that the software can be improved continually without the users worrying about software updates [Kreckwitz, 2007].

Although TM architecture models do not seem to attract any interest from the TM research community, the processes that lie within the TM system's internal operation, especially the ones that involve some kind of Natural Language Processing (NLP), have not stopped drawing the attention of the vast majority of TM researchers since the early days of the system. In fact, a general overview of TM-related literature reveals a degree of hype around research efforts concentrated on the improvement of the key TM processes, with new approaches being proposed one after the other, in a constant quest for optimum solutions to thorny NLP issues.

1.3.4. Key internal TM processes

Before discussing different approaches to TM processes, it is important to describe briefly the internal workings of a basic TM system. In a typical TM workflow scenario, the user (i.e. normally a translator) starts by opening his source document in a text-processing environment, either the one provided by the TM system or the one that the system ports to (e.g. Microsoft Word). Due to the fact that documents to be translated can be in any file format, it is essential that the system includes file format filters, so that it can convert the proprietary format into a format that the system can read.

1.3.4.1 Segmentation

Once the conversion of file formats is complete, the system starts the *segmentation* of the source text. In essence, it splits the source text into units ('segments'), which can be of any length (e.g. words, phrases, sentences, paragraphs) depending on the segmentation strategy of each system. Usually, the punctuation marks (e.g. periods, exclamation marks and question marks), hard returns and text formatting indicators (e.g. paragraph marks, list tags) determine the boundaries of each segment, but the user is also allowed to customise the segmentation rules, for example, by adding stop lists (i.e. by specifying words or abbreviations, such as "Prof.", which should not be considered as segment separators). However, accurate segmentation, or as often called 'sentence boundary disambiguation' [Mikheev, 2003:212], is a challenging task, because each language has its own script, so that the principles that drive how a sentence is constructed and delimited differ from one language to another [Savourel, 2001: 209]. Similarly, if segmentation is attempted at a finer level of granularity (e.g. words), an additional set of problems arises from the way the words are delimited in each language. For instance, in English, if a token is surrounded by spaces, tabs, newline characters or punctuation marks, it can be treated as a distinct word,

but this does not necessarily apply for other languages. The presence of compounds, word contractions and numbers written in full can also create problems for a tokeniser [Farghaly, 2003:225-7].

Segmentation is the first key TM process, as it impacts directly on the quantity and quality of matches retrieved by the system in the subsequent process of matching. For instance, if the segmentation is done at paragraph level, the system has little chance of finding a match for the source paragraph in the repository (a phenomenon called 'silence' by NLP researchers). However, if it does find a paragraph match, then this typically does not require any editing and can be incorporated in the translation as it is [Esselink, 2000:363]. In the next case, where the system performs sentence-based segmentation (full sentences or clauses), there are more chances of it finding a match from a repository that contains sentences. Whether those matches are correct or not, it depends on the match retrieval mechanism of each system. Finally, in the case where a system segments the text in smaller units (e.g. phrases, words), there are even more chances of it finding matches [Somers and Fernández Díaz, 2004] – as a matter of fact, so many that the system produces the phenomenon of 'noise' – but naturally most of them are useless in the context of the source text [Simard, 2003b]. In such an event, the system must be able to spot the correct matches by applying a match filtering mechanism.

For the last ten years, many methods have been proposed for the segmentation of texts by researchers in both TM and MT. In recognition of the fact that whole sentences rarely repeat in a document to be translated (unless it is a manual), and based on the premise that long segments (e.g. full sentences) reduce the chances of the TM system finding matches [Gervais, 2002; Hunt, 2003], research has been geared in favour of sub-sentence segmentation [see Varea et al., 2005; Gotti et al., 2005; Casillas et al., 2004; Simard 2003; Huang et al., 2003; Langlais and Simard, 2001; Macklovitch and Russel; 2000]. A major distinction between the proposed methods involves the contrast of approaches based exclusively on the information contained in the text to be segmented (e.g. punctuation marks, spaces, formatting indicators) along with statistical calculations [see Varea et al., 2005; Macklovitch et al., 2000], approach followed by virtually all commercial TM systems [Garcia and Stevenson, 2006:1], and those approaches that depend on complementary linguistic knowledge. This can either be extracted from the translation memory (or the terminology databases) and used to train the system on how to do efficient segmentation (data-driven approaches) [see Chenon, 2004; Callison-Burch et al., 2004; Reinke, 1999], or it can be linguistic knowledge that has been introduced into the system (rule-driven approaches) in the form of pre-determined linguistic rules or phrasal lexica [see Gotti et al., 2005; Planas, 2005; Gough et al. 2002; Schäler, 1996]. A critical review of these approaches, as well as a description of corresponding experiments, fall beyond the scope of this thesis.

1.3.4.2 Data storage

Most TM systems segment the source text before storing it in the database. Each segment is accompanied by contextual information, such as the name of the client, the domain, date and name of translator. However, some systems (such as MultiTrans and LogiTrans) follow a different approach, called the 'full-text approach'. Instead of segmenting the texts at the beginning of translation, they store both source text and its translation as full bitexts using the character-string-in-bitext (CSB) technique [Gow, 2004]. Once the bitexts are in the database, they are aligned at paragraph level. This approach has two main advantages compared to the conventional method: a) the faster creation of a large TM database containing previously translated material and b) the retention of the co-text for any match found and suggested to the user [Gervais, 2002].

There are also systems that do not require the use of a database for the storage of the previously translated material. STAR Transit, for example, instead of creating a physical TM database, builds and makes use of a 'virtual' translation memory, which is essentially an index of selected reference material residing in any directory on one's computer. The user chooses the reference material he wants to use (i.e. the translated documents that are relevant to his project), and the system, after extracting the text into XML files, creates the index with the associated files. This idea has many advantages over the classic database. Firstly, because the index is resident in RAM, the searches for matches are relatively quicker. Secondly, the user has greater freedom in tailoring his repository according to the needs of each project. In addition, this model of a TM repository works especially well for projects with many updates containing a lot of small changes [Zerfaß, 2002], although the greatest advantage of this idea is perhaps the absence of issues related to database maintenance and to the security of data residing in a database (databases are, by nature, prone to crashing or getting corrupted).

1.3.4.3 Match search and retrieval

Returning to the typical TM workflow, after the system has segmented the source text, the translator chooses the segment he wants to start translating and the system automatically looks up this segment in the TM repository. When a match is found, the system retrieves it and suggests it to the translator as a possible translation.

Matching is the second key TM process, as it determines the ability of the system to leverage phrases and terms from existing translations in an effective and efficient way. According to Bowker [2002:95-105], there are several kinds of matches: *exact matches*, *fuzzy matches*, *full matches*, *sub-segment matches* and *term matches*. Yet, recently, some TM systems have been able to produce an additional kind of match: an *MT match* [O'Brien, 2006:187-8]. A brief description of each match type is provided in Table 3.

Type of match	Description
exact (or perfect) match	The segment in the TM is 100% identical to the source segment in terms of morphology, syntax and sometimes formatting (in other words, it is the same in spelling, punctuation, inflection and numbers). The process followed by the TM system to recognise exact matches is simple string pattern recognition.
full match	The segment in the TM differs from the source segment only in terms of variable elements, such as numbers, dates, times, currencies, measurements, and sometimes proper names.
fuzzy match	The segment in the TM resembles approximately or partially the source segment. The differences between the two segments are usually highlighted by the system. Fuzzy matches are presented according to a similarity degree given by the system. The user usually specifies a threshold for the acceptable similarity degrees.
sub-segment match	The source segment is identical with part of a segment in the TM.
term match	The source segment is identical with an entry in the term base (or the lexicon database).
MT match	When the source segment does not match with any segment in the TM, the system constructs a segment usually by combining sub-segment matches or by generating a new match by means of Machine Translation techniques.

Table 3: Types of TM matches

Issues in match search and retrieval (due, for instance, to the inflectional and derivational morphology of words) have been widely discussed [see Bowker, 2002: 72-4; Somers, 2003:37-41] and several matching techniques have been proposed during the past few years [see Hodász and Pohl, 2005; Järvelin and Pirkola, 2005; Vertan and Martin, 2005; Kraaij, 2003; Simard 2003; Kamps et al., 2003]. All matching techniques aim at rendering the TM system efficient not only in retrieving *all available* exact or fuzzy matches for a source segment (match recall), but also in retrieving the *correct* exact or fuzzy matches (match precision).

The two match retrieval techniques employed by current commercial TM systems are: a) segment-based matching and b) character-string-based (CSB) matching [Gow, 2003:22-44]. Each of the two techniques is directly related to the way in which the texts are stored in the database. More specifically, a system that segments the text in translation units before storing it in the database is normally using segment-based matching, whereas a system that does not segment the text, but stores it as full text instead, performs a CSB matching. Most

TM systems (e.g. SDL Trados, Wordfast) implement the first technique, which means that they look for matches in the sequence of character strings of each segment in order to identify potential translation equivalents. The similarity between segments is operationalised via the distributed similarity of each word in the segment [Hunt, 2003; Simard, 2003]. Some TM systems are also able to recognise matches not only at segment level but also in sub-parts of the segments that exist in the TM database (e.g. Déjà Vu X). Alternatively, systems such as MultiTrans and Lingotek use a matching algorithm that calculates the similarity of equivalent continuous character strings in a text, rather than relying on segments and fuzzy matching [Garcia and Stevenson, 2006:24]. Both these statistical techniques are based on character n-gramming and are language-independent.

However, recently, TM systems of the 'second generation' have opted to use matching techniques that are linguistically enhanced. These can be lexeme-based or grammatical pattern-based techniques [Planas, 2005, Gotti et al., 2005]. Based on the premise that superficial matching techniques alone may result in a potentially useful match being overlooked, these systems make use of morphological and grammatical information to improve the match coverage. A more detailed description of the process followed by each second-generation TM system is given in section 1.3.5.

According to experiments conducted by Carl and Hansen [1999], when comparing matching results from a string-based TM, a lexeme-based TM and an EBMT system, the least generalising system (the string-based TM) achieved higher translation precision when near matches existed in the database. However, when the database did not contain any similar translation examples, the lexeme-based TM performed better in terms of translation coverage but it lost in precision (the EBMT system outperformed both in match recall but performed poorly in match precision). Nevertheless, these results were based on a fully automatic evaluation method; hence, they could not establish the superiority of a particular matching technique in real work conditions. Such an evaluation that includes humans is very important as translators are able to assess not only if matches are indeed the *correct* translations for a source segment indeed, but also if they are *suitable* for the context of the source text.

From a general perspective, the main disadvantage of TM systems that use linguistically enhanced matching techniques is that they are language dependent (since they rely on the built-in language resources). Consequently, such systems can work only for a small number of language combinations for which adequate language resources have been developed and incorporated into the system.

1.3.4.4 Match assembly

In the past two decades, several attempts have been made to combine TM systems with MT ones, to compensate for the limitations of the first and boost the TM system's productivity.

From a cursory look at the translation software market, it is quite common to see TM systems with the ability to port to a MT system and vice versa. A successful example of such an integration of TM and MT is the EURAMIS system used by the European Commission since 1995 [European Commission DGT, 2007]. Translators who use this system are offered various match retrieval options, including the possibility for a match request from an integrated machine translation system:

> When Euramis does not find any perfect or fuzzy matches in the central memory, it requests a machine translation, which is included at the end of the retrieval result. In [TRADOS Translator's Workbench...], a machine-translated sentence of this kind is displayed in grey to distinguish it from perfect (green) or fuzzy (yellow) matches. The advantages compared to retrieval or machine translation alone are obvious: users always get a suggestion while being aware of its status, and they do not have to keep making the same changes in machine translation, since validated proposals from machine translation become perfect or fuzzy matches respectively if they reappear later in the text. [European Commission DGT, 2007:18]

The idea of bringing machine translation closer to TM systems has also existed for a long time in the realm of academic research as a possible way of expanding the capabilities of TM technology and of exploiting fully the available resources in the TM repository [see Carl, 2000; Carl et al., 2002; Simard, 2003; Huang et al., 2003]. Gradually, TM developers started investigating how a TM system could employ MT techniques (instead of porting an MT system to a TM system), so that it could suggest a match, in an autonomous way, when the TM repository did not contain an identical or similar TL segment. It did not take long before MT techniques found their way into TM systems.

The first commercial TM system known to have implemented EBMT techniques in a dynamic sequential manner, in order to be able to construct a match, is DéjàVu X.[24] When the system cannot find an identical or similar match in the segments of the TM repository using string-based fuzzy techniques, it then looks at the parts of each segment. When for a source segment the system finds two sub-segments existing in two different segments stored in the TM database, it puts them together to form a new segment that will be suggested as a match (the feature is called 'Assemble from portions'). The system can also carry out replacements of both invariable parts (numbers, proper nouns) and variable parts with declensions. However, this is a mere 'find and replace' process, which means that the new match has little chance of being a grammatically correct segment, especially if it has been generated from a highly inflected language.

This capability of the system draws on three essential prerequisites: a) that the system is able to identify matches at the sub-sentence level (as discussed in section 1.3.4.3), b) that it is able to score multiple fragments found as a translation for a portion of the source segment, and pick the best and c) that a combination algorithm is employed for inserting the fragments in the correct place within the match that is going to be proposed. The

[24] All information about Déjà Vu X has been drawn from the Déjà Vu X Professional Users' Guide (available from Atril's website: http://www.atril.com/).

second point relates to the issue of overlapping translation examples contained in a TM database, as identified by Somers:

> Some examples will mutually reinforce each other, either by being identical, or by exemplifying the same translation phenomenon. But other examples will be in conflict: the same or similar phrase in one language may have two different translations for no other reason than inconsistency. [1999: 121]

Furthermore, this issue becomes even more complex when the system is confronted with word ambiguity (or simply put, words that have many meanings, and which might therefore have many translations). Since no semantic analysis is performed, the system must rely on its statistical scoring system, in order to decide which target fragment to use.

On the positive side, the match assembly capability of the system is further enhanced by the comparison of source segments not only to the TM and termbase entries, but also to a third database containing word lists called 'lexicon', thus exploiting more information. This helps, also, in building more complete target segments, since the 'lexicon' normally contains general purpose words like 'and', 'or', 'from', etc. Obviously, in order for the system to be able to suggest useful combinations of fragments from across databases, it is essential to have a sufficient quantity of entries stored in each database.

Déjà Vu X's example of integrating MT techniques into a TM system can be seen as an early implementation of the attempt to combine the strengths of the two technologies in a seamless way for the benefit of translation professionals. While automation is attempted in a process reserved for the human intellect (i.e. text generation), the function is offered purely for the purpose of assistance and comes as an option, which may potentially add value without compromising the control of the user over the system. What is more important, Déjà Vu X does not make use of any external linguistic information, which means that the system offers the same matching function irrespective of the language combination.

It is interesting to note here the fact that Déjà Vu X chose to implement EBMT techniques instead of any other MT technique, namely rule-based or statistical. One possible explanation may be that an EBMT technique (just as a statistical technique) can work without the appendage of external language resources. It is also widely acknowledged that EBMT techniques could fit most appropriately in a TM model. Such an assumption derives from the similarities between TM and EBMT systems, clearly illustrated by Somers and Fernández Díaz [2004], like the availability of a corpus of aligned translation examples in both systems and the common processes of segmentation, matching and alignment. Just like most TM systems, some EBMT systems also perform matching by comparing sequences of character strings [Somers, 2003:514], which makes a possible convergence of the two systems clearly very sensible.

Another approach to match construction is the one followed by 'second generation' TM systems. Unlike the minimalist approach of Déjà Vu X, some TM systems have sought to exploit the TM databases based not only on the bare textual contents but on the structural and syntactic contents of segments. To exploit a database in such a manner, however, requires the introduction of external linguistic knowledge that will help the system identify these structures and allow for a more sophisticated processing. A description of the match assembly capabilities of 'second generation' TM systems is offered in section 1.3.5.

Returning to the typical TM workflow, after the system has been able to offer a match, the translator has then to decide whether to reuse the translation example, to adapt it or to ignore it and insert his own translation. Once the segment is translated, the system automatically saves the pair of source and target segments in the repository. As the translator works his way though the source text, each successive segment is looked up automatically in the TM repository, and a translation is proposed (if one is available). By the end of the translation of the source text, the system has stored all pairs of source and target segments in the TM database for future reuse. The following illustration (Figure 3) shows how all the basic TM processes are linked together in a typical TM workflow.

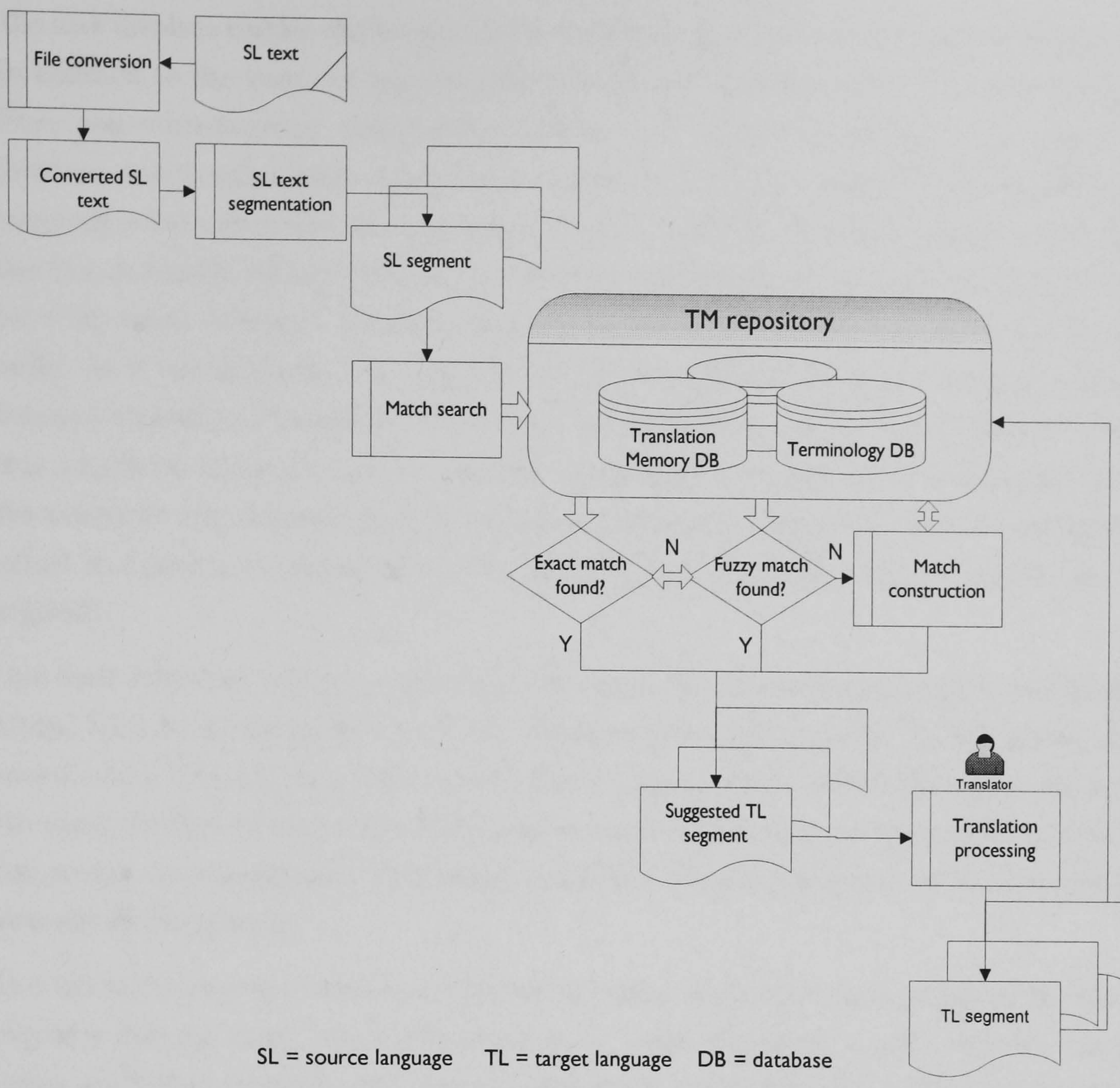


Figure 3: Basic processes in a simplified TM workflow

1.3.4.5 Alignment

TM systems usually come with no, or virtually empty, repositories. According to the ICL TM Survey 2006 [Lagoudaki, 2006], the most frequent way followed by translators to build a repository, is to create a database and have it filled as they translate (by storing pairs of segments, as described previously) (*interactive method*). The next most frequent way is to build up a repository with aligned original texts and their translation (*retrospective method*). *Alignment* is a feature offered by most TM systems and it constitutes an additional key process within a TM system. In fact, it impacts directly on the effectiveness of the matching, because if a source segment is not accurately aligned with its translation equivalent, a wrong match is likely to be retrieved. Before the alignment process, the TM system normally segments both source and target texts and then tries to establish correspondences between their segments.[25]

[25] TM systems that follow the 'full-text' approach do not segment the texts prior to alignment and storage. Instead, they store the texts in full and then they align them at paragraph level (see section 1.3.4.2).

The task involves certain challenges that inevitably compromise the accuracy of alignment. In addition to the issues of segmentation (mentioned in section 1.3.4.1) usually deriving from non-corresponding punctuation systems and a different notion of a 'sentence' [Somers, 2003b:35], a single source text sentence can often be translated by multiple target language sentences or vice versa, or information can be omitted from or added to the target text (e.g. to handle cultural references). Moreover, sometimes source and target texts do not have the same structure, because sentences or whole paragraphs may be in a different order. As a consequence, it is not always easy to establish one-to-one correspondences between sentences. Thankfully, TM systems are interactive tools by nature, which means that automatic alignment can be corrected any time by the user; either immediately after the automatic alignment process, by joining or splitting misaligned segments, or during the actual translation, by accessing the TM database and by editing the appropriate pair of segments.

One basic statistical method of automatic alignment was first proposed by Gale and Church [1993] who, by assuming that there is a significant correlation in the relative length of a source and a target segment, anticipated that a target segment would have approximately the same number of characters as the source segment, in order for the system to draw a one-to-one correspondence. This simple statistical comparison seems to be followed by virtually all TM systems.

In order to improve the accuracy of this method, other researchers have proposed the use of cognates [Simard et al., 1992; Johansson et al., 1993; Melamed, 1996]. Cognates can be either words that appear in the same (or similar) form in both source and target languages (irrespectively of their meaning), or words like proper names, dates, abbreviations and acronyms, or even punctuation marks and numbers [Somers, 2003b:37]. Those cognates serve as anchor points that can help a system link a source and a target segment more accurately.

Alignment can be performed on various levels, depending on the segmentation output. In other words, if a system segments the source text at a sentence-level only, it will then try to align sentences; if it segments the text in 'chunks' (i.e. smaller units, like phrases), it will need to align the 'chunks'. In the latter case, the task of alignment becomes even more challenging because the system needs to identify the similarity between the two 'chunks', and a mere string-length comparison cannot be much of a help. To this end, an appropriate morpho-syntactic parsing schema and/or detailed tagging is used to identify any similarities at grammatical and syntactic levels. Examples of TM systems that perform alignment at a sub-sentence level are SIMILIS [Planas, 2005] and MetaMorpho [Hodász and Pohl, 2005].

BIBL LONDON UNIV

1.3.5 TM system generations

From the perspective of computational linguistics TM systems qualify as NLP systems, since they deal with natural language which they are required to process to some degree. According to Nirenburg et al. [1992:117-8], there are four levels of language processing that NLP systems aim to able to perform: morphological, syntactic, semantic and pragmatic. If we draw an analogy to the level of technical sophistication of TM systems, we could define two, or perhaps three, generations of TM systems.

The *first generation* of TM systems is hardly capable of any language processing, with the exception of few TM systems (e.g. Déjà Vu) which are able to perform some basic morphological processing ('shallow processing') that does not require any additional linguistic knowledge [Benis, 2003a:24]. The vast majority of TM systems currently available on the market belong to the first generation of TM systems.

The *second generation* of TM systems requires the introduction of linguistic knowledge into the system in order to be able to perform morpho-syntactic analysis on the source and target texts. Two commercial TM systems are known so far to belong to this generation: SIMILIS[26] and Masterin[27]. The two systems are described briefly below, forming a picture of the state of the art in TM technology.

SIMILIS, in general, applies linguistic rules to a number of processes, including segmentation, alignment and automatic extraction of terms and phrases from translation memory content[28]. After segmenting the source and target texts at sentence level, it runs a linguistic analysis and further splits each sentence into syntactic units ('chunks') attaching grammatical annotations to them (this is performed with the help of monolingual lexicons and algorithms that can recognise grammatical categories) [Planas, 2005]. It then indexes those as translation units as well. So, every time the system searches for a match, it looks not only at the sentences, but also at the chunks (thus increasing the possibilities of finding a match), and especially those chunks that are in the same grammatical category as the source segment (thus increasing the possibility of finding the right one).

Masterin, on the other hand, segments the source and target texts in a flexible way according to the examples available in the TM database ('Knowledge Base') provided. Each segment is annotated with grammatical information (with the help of a POS tagger) and constitutes a grammatical 'translation pattern'. So, matches are sought by a deep-structure pattern recognition method that looks beyond the surface appearance of segments. If several matches are found, the system determines the best match by using semantics (with the help of a built-in lexicon) and/or examining their use frequency or domain information. In the case where no match is found, the system constructs and suggests a fuzzy match from

---

[26] Developed by Lingua et Machina.
[27] Developed by Master's Innovations.
[28] Based on Jost Zetzsche's newsletter of 23/02/2007 (82nd Tool Kit).

the available resources in the database by applying translation heuristics [Grönroos and Becks, 2005].

From an overall examination of the literature on second generation TM systems no evidence exists for any human evaluation of the systems. Therefore, it is impossible to determine if any of the systems outperforms the rest or, most importantly, if the systems of the second generation surpass the systems of the first generation in usefulness. In contrast, what is very apparent from the relevant literature is the excessive insistence on the technical sophistication of the system, without this being associated with practical benefits for its users.

Perhaps prematurely, since no actual proof has established yet that second generation TM systems have succeeded in their purpose, a *third generation* of systems has been proposed recently. Pekar and Mitkov [2007] suggest that it is not enough for a TM system to perform morpho-syntactic analysis on the source and target texts, but it also needs to perform semantic analysis that will help the search engine to have a better idea of what the segments are "about", so that it can retrieve more relevant matches. Their proposal consists of a procedure that considers the syntactic, lexical and lexico-syntactic variability of segments and is interested in arriving at generalised semantic representations (with the help of the WordNet lexical resources) that can be attributed to each segment. The retrieval mechanism operates on these generalised representations in order to determine useful translations for a given source segment. In their most recent work [Mitkov and Corpas, 2008], they concentrate their efforts in the identification of rhetorical predicates for each TM segment with the belief that sentences labelled with the same rhetorical predicates have a greater probability of being semantically equivalent. These latest techniques have not been implemented in an existing TM system so far, therefore no human evaluation has yet been able to determine their usefulness for the translator.

1.4 Social context of TM systems

So far, TM systems have been looked at from a short focal distance to provide an understanding of the technology in its technical dimensions. However, technology is not created in a vacuum. It exists in a social context, which defines its usefulness and proves its significance for society. This context is formed by a constellation of relationships between stakeholders and the TM system, and between stakeholders themselves. TM systems have two primary stakeholders that determine the purpose of the system: the language services industry (who are also the clients of the TM software developers, as well as the indirect users of the system) and the translation professionals (who are the end-users of the system, but may not be the clients of a TM software developer if the buying decision has been taken by their employer). It is important to analyse this context in order to understand the most

important factors that drive the development of TM systems in a business-minded world. Do these factors relate to the needs of translation professionals or to those of the language services industry? What is the difference between the two sets of needs and why does one set dominate over the other?

1.4.1 The language services industry

The language services industry which includes companies in both the translation and localisation sectors has endured dramatic changes in terms of its needs during the past few decades. These have been induced by a combination of technological, political and social events.

On one hand, the growing penetration and use of the internet, along with technological advances in information systems and telecommunications have turned globalisation into a reality. This practically means that, at least in the developed world, people and businesses are increasingly acquiring the means to communicate with each other, irrespective of their physical location and language. But no communication can be effective if people are speaking different languages. For this reason, translation has become a necessity that grows as the digital divide gets smaller.

On the other hand, the growing exploitation of the web as a commerce and marketing platform for millions of multilingual consumers has amplified the need for translation and has given birth to a more specialised service that involves translation: the localisation of content and products. Even though English remains the commercial lingua franca, its hegemony is clearly eroding as more users from all over the world obtain access to the internet. The explosion of web content addressed to an increasingly diverse multilingual group of people has highlighted the need for multilingual content delivery. Similarly, in order to meet the demands of a global market, to satisfy customer requirements, to seize opportunities in new markets, and even to gain the advantage of differentiation among their competitors, product manufacturers have resorted to localisation, which they increasingly see as a strategy for business growth.

Major political decisions have also had an impact on the demand for translation. Since 9/11 and the expansion of the European Union, government-mandated translation has risen significantly. In addition, European regulations that want all product documentation translated in all official EU languages have rendered translation not only a legal requirement, but also a political obligation.

Moreover, a further rise in the demand for translation has been encouraged by social trends in the last decade. In the spirit of social corporate responsibility, an increasing number of businesses, especially in the technology sector, attach social value to the effort of reaching people in developing countries. In order to make their products socially inclusive they turn

to localisation which helps them adapt the products linguistically and make them fit for the particular cultural context. In a similar spirit of social inclusiveness, several initiatives for localising open-source software or free content have emerged on the web, in order to encourage accessibility to technology and information. The impact of these initiatives might not be significant for the language services at present, because they are few, but as they expand in a limitless global platform of information such as the web, one should expect that they will put pressure in the language services industry in the future. In general, multilingualism is increasingly promoted as a social policy, especially by NGOs, governments and the European Union, which means that the social trend of localising content and products can only be amplified in the years to come.

According to Common Sense Advisory, a US-based business consulting firm, the language services market experienced an unprecedented growth with an estimated total of US$9.5 billion in revenue in the worldwide translation and localisation services market in 2006[29]. Their analysis reports an average projected growth rate of 7.5% per year for the next five years, which is undoubtedly an encouraging forecast and suggests even greater demand for translations to come in the near future.

However, the exponential rise in the volume of translations has brought about a number of challenges for the language services industry, some of which are the following:

1. **pressure to deliver high quality translations in shorter turnaround times**:

   Digital content, especially webpages, have a short life span because they are updated frequently. This is facilitated by Content Management Systems and other web design tools which have rendered the content editing easier and the publishing instant. For businesses that maintain a presence on the web, as well as for news broadcasters, it is important that multilingual information be fresh at all times. Therefore, translation needs to be done as quickly as possible. Similarly, when multilingual documentation accompanies a product that has a launch deadline, this deadline will apply for translation too. Frequently, the translation of the documentation comes too late in the product development lifecycle, and as a result, translators are on tight deadlines to avoid holding up the release of the product.

2. **requirement to handle digital content that derives from a variety of new media sources:**

   Translation requests do not concern only print material and webpages. Text may be contained in a variety of new media packages, such as audio, video, films, DVDs, and graphics.

[29] The full report by Common Sense Advisory can be read at http://www.commonsenseadvisory.com/members/res_cgi.php/070502_Q_Top_20.pdf [last accessed on 22/07/2008].

3. **increased cost of human translation:**

As discussed earlier, translation demand is on the rise, but the same is not true for the numbers of professionals who can produce high-quality translations. In fact, as many industry commentators have observed, there is a shortage of skilled technical translators, [McCallum, 2004:27] who could bear the weight of the enormous volume of translations without any help. Because of this discrepancy in levels of demand and supply, it is natural that translation costs have risen. This phenomenon is particularly acute in the supply of translation services for the biosciences sector, where demand outgrows delivery capacity.

Schäler [2002], in his analysis of the problems of the language services industry reveals – contrary to one's expectations – the industry's lack of profitability because of inadequate organisational structures being unable to cope with the new challenges. He argues that the only way forward for the industry is the transformation of the established supply chains through technology. Today, it is widely acknowledged that the vast majority of language service providers have realised the potential of translation technology as the solution to their concerns and have taken major steps in integrating it into their workflows.

1.4.1.1 Significance of TM systems

TM systems represent one of the available translation technology solutions that can be used by language service providers. Others include terminology management systems, MT systems and, lately, translation management systems. Yet, TM systems have proved to be more successful than the rest in their role of assisting the delivery of high-quality translations.

Several commentators have discussed the advantages of TM technology for the language services industry and have provided evidence of tangible benefits. Esselink [2000: 362-6], in his overview of the importance of TM systems for localisation, argues that such benefits relate to:

- **Productivity increase**

Productivity levels can be improved by approximately 30% or even 50% [Esselink, 2000: 362-6]. This is achieved mainly for four reasons: a) a translation that exists in the TM repository will be used again and again when the same source text reappears for the translator, thus saving the time and effort that would be spent on translating anew the same source text every time it appears; b) TM systems provide facilities (e.g. statistics on word counts and segment repetition within a translation project) which offer valuable information for project management and scheduling; c) several translators can work on the same project in a networked environment for speeding up the completion of a translation

project; and d) TM systems with a variety of file format conversion filters make it possible for language service providers to deal with content deriving from a variety of media sources.

- **Translation quality assurance**

TM systems can improve quality mainly in two ways: a) terms, brand names and standard expressions that have been translated before and are stored in the database are used consistently every time they appear in a project of the same domain or for the same client; and b) every TM system offers Quality Assurance facilities that help translators (or proof-readers) track any inconsistencies between the source and target texts, thus alerting them to mistakes and/or omissions in the final translation.

- **Costs reduction**

Reduction in costs may result from various TM advantages, such as a) the recycling of content, which eliminates the need to translate a text/paragraph that has been translated before and is available in the TM repository and b) the fact that most TM systems normally come equipped with file format converters, saving language service providers from buying the specialised software that would allow them to process a proprietary format file containing the text for translation.

Based on the above benefits testified through numerous case studies published regularly by TM developers[30], TM systems seem to be of paramount importance for the viability and growth of the language services industry in the current socio-techno-economic conditions.

It is important to notice here that TM systems seem to address successfully some of the concerns of the language services industry, as those were mentioned earlier.

1.4.2 The translation professional

The same circumstances that impacted on the language services industry affected inevitably the translation professional, as the latter constitutes the central figure in the translation supply sphere. Several researchers like Bowker [2002:11-13] and Austermühl [2001:1-8] discuss at length how the socio-techno-economic conditions have impacted on translators' work, but limited mention is made of the concerns that the current conditions have raised:

1. **pressure to produce high quality translations within shorter deadlines**

   Concern shared with the language services industry.

[30] On almost every TM developer's site we can find case studies on the implementation of TM systems by their clients. Examples of case studies can be found at: http://www.sdl.com/localization-information/case-studies.htm [last accessed on 12/09/2008].

2. **requirement to handle digital content that derives from a variety of new media sources**

   Concern shared with the language services industry.

3. **balance of excellence in profession with job satisfaction**

   This is a concern that all professionals have regardless of the industry they belong to. It is an important concern because it has implications for the productivity of the translator and it needs to be considered seriously in the design of environments or tools for translators. It is also a concern that is hardly measurable and can vary from person to person; it is therefore difficult to address.

The above concerns could give an *a priori* indication of the needs of modern translation professionals, which this thesis plans to identify in a systematic way in the following chapters.

The general consensus among researchers in Translation Studies is that the use of technology has become a necessity for the modern translator in order to stay competitive, or even to survive in the translation supply market [Austermühl, 2001:7-8, Bowker, 2002: 13-14; O'Hagan and Ashworth, 2002:xi]. In these conditions, TM systems have been proclaimed as a solution to the translators' concerns not only by TM developers but also by many university translator training programmes (e.g. at Imperial College London, Dublin City University, University of Limerick).

1.4.2.1 Usefulness of TM systems

TM systems are used by translation professionals as tools, therefore their potential is measured in terms of usefulness. The potential of TM systems for translation professionals have been discussed by many TM researchers, such as Ray and Ray [1999], Samuelsson-Brown [2004:77-8] and Bowker [2002: 114-7]. In summary, such potential benefits relate to:

- Productivity increase (more translations in greater speed, ability to process a variety of source file types).
- Improvement in the quality of translation output (mainly due to consistent use of terminology).
- Reduction of repetitive work, thus enhancement of job satisfaction.

These alleged benefits of TM systems for the translation professional have been contested by several researchers, mainly due to the non-existence of related epistemic proof deriving from research on translators' perceptions and work practices around TM systems. Pym [2006], for example, challenges the overall benefit of TM systems when he argues that the

technology – in its present form – separates the communicator (i.e. the translator) from the act of communication. He describes a situation where:

> Translators usually cannot see the whole source text; they are unaware of any iconic qualities that might profile a particular end-user; they are normally given no specific information about the purposes of the communication act; [...] they rarely receive any feedback from the distant world of text use; and if any of those things happen, it is at the stage of revising, publishing or marketing, beyond the technologies that keep the translator looking at nothing but phrases from memories.

In an effort to investigate translators' perceptions around TM systems, Dillon and Fraser [2007] examined the attitudes of users and non-users of TM systems towards TM systems. They analysed correlations between a positive or negative attitude and specific categories of translators (as those were determined by e.g. experience and IT skills) and mentioned the fact that there are unclear perceptions around benefits and limitations of TM systems, but they did not provide any proof of certain benefits deriving from the use of TM systems, nor did they analyse the perceptions of TM users in relation to the usefulness of their tools.

The TM Survey 2006 [Lagoudaki, 2006] has also attempted to shed light on TM users' perceptions and has revealed that TM users have indeed realised benefits such as savings in time, improvements in translation quality (through terminology use consistency) and sharing of language resources. However, those perceived benefits are not contrasted with any perceived inconveniences deriving from the use of TM (no such information is requested from participants) and are not analysed, so that the survey cannot provide sufficient proof of the cumulative benefit gained by translation professionals by the use of TM systems.

Earlier empirical work [e.g. by Webb, 1998] cannot be considered to offer any evidence of TM benefits for translation professionals, as the TM systems in reference to which perceptions were analysed are far behind in design and technology from present TM systems.

The question of whether the current TM systems truly benefit translation professionals thus remains open. Any further discussion on this issue goes beyond the scope of this thesis. What is of interest to us is that there is serious scepticism around the actual usefulness of TM technology for its users, with regards to existing TM products.

1.4.3 What drives TM system development

By comparing the obvious needs of the language services industry and those of translation professionals, we notice many overlaps. Both groups feel the pressure to deliver high-quality translations in shorter timescales and to respond to requests made by their clients. The difference appears when the needs derive from different goals that each group is set to achieve. For instance, while the industry is interested in reducing the financial cost of the

translation production, the translation professional does not have such a concern, because, apart from overhead costs (such as cost of training and investment in technology), no significant financial costs normally occur from his work. For translators, human factors are important, instead, such as the quality of the environment in which they work and the ergonomics of the tools that they find useful, both of which have direct impact on the level of job satisfaction obtained.

In a business-oriented world, where system development is driven by profit potential, it is no surprise to see that TM systems have been developed to serve the needs of the language services industry, rather than those of translation professionals as individuals. The industry's effort in not only containing, but also reducing service costs has driven the hasty development of translation tools without giving the necessary attention to the end users.

The language services industry is undoubtedly a large interest group with strong financial and political power enabling it to drive the development of technology. Translation professionals, on the other hand, have never been able to form a powerful entity that can further the interests of the profession. Therefore, translators have been forced to use the technology that was designed to serve industry goals and not their needs.

TM systems' development is driven by market forces, by capabilities afforded by software companies and computer systems and by the clients who dictate what they want from language service providers. Every effort in TM systems R&D must always be in line with commercial considerations. With the rare exception of free software, a TM system has a monetary value, therefore it makes business sense that priority is given to the needs of business buyers (i.e. translation/localisation companies), rather the needs of consumer buyers (i.e. translation professionals). Every investment in R&D has a cost and a risk, therefore trade-offs are more likely to satisfy the first group.

Having said that, no one can argue that no real consideration is given to translators' needs by the TM software developers. TM developers are making considerable efforts in eliciting information from the end-users of their products (e.g. through online user groups and product feedback forms). However, these efforts prove insufficient, when we can still see system design - user needs misalignments, instantiated by complex and impractical systems. It is obvious beyond any doubt that many of the needs of modern translation professionals remain unmet by the existing TM systems. This is not only because TM developers may fail to identify them, or may downgrade them as a result of business priorities, but also because needs change constantly and at the same pace as changes happen in the world around us. Serving a societal need is not enough for any technology. It quickly saturates the potential of the system. In contrast, user needs can offer endless opportunities for expanding the possibilities of a system. It is the present research's goal to identify these needs, and analyse them taking into account the diversity in user characteristics, tasks and TM use environments.

Chapter 2: Preparing for user needs investigation

There is an abundance of evidence in the software engineering literature that proves that software quality is inherently linked to the correct and early identification of user needs [Xia, 2003; Ryan, 2000]. A widely acknowledged premise is that quality is achieved when the system successfully meets the needs of its users, thus creating user satisfaction. Although the importance of needs identification is commonly recognised, many software development projects are reported to undervalue the task, or fail to correctly identify needs with fatal consequences to the entire project [Townsend, 2007: 9; Kohler and Paech, 2002]. This failure is commonly associated with hasty activities during needs identification, lack of resources to spend on the task and lack of commitment or managerial support. Indeed, needs identification is considered as a costly process that can take a significant part of the product planning schedule and it is likely to bring delays to the software project. This is the main reason why only big companies can afford to perform this task properly [Griffin, 1993:7-8]. Usually, most of the challenges derive from the fact that user needs identification is a collaborative activity that requires the end-user involvement in the software development process. Nevertheless, any lapse in proper identification of user needs may result in the incorrect specification of the system, which can also require substantial resources and time in order to fix the software's defects [Eberlein and do Prado Leite, 2002].

Investigating user needs might seem like a simple capture of knowledge from the user domain, but it is far from being a straightforward task. According to Gotel and Finkelstein [1994] the generic sources of complexity in such a task can be the following:

- A stereotypical end-user cannot be predefined. The needs of users will differ and will often be inconsistent.
- Users tend to change their minds often about their needs.
- The quantity, heterogeneity, and depth of detail of the potential information required, precludes predefinition.
- It is difficult to predefine how any access to information, and its subsequent presentation, will be required.
- The data acquisition relies on personal contact, as there is always something out of date, undocumented, inaccessible or unusable.
- Not all users perform the same tasks on a system. Each user group, which represents a different end-use context, exhibits unique needs which may be in conflict with the needs of users in other use contexts.

A successful investigation of user needs must, therefore, consist of a chain of activities, from information gathering to dissemination, which take into account all the above problems and address them in the most suitable way. It needs to be systematic, yet creative, and well prepared. The line of investigation that sets the general framework for the activities to be carried out in the present research is described and justified in the next section.

2.1 Line of investigation

Due to the absence of previous research on user needs identification for translator-support systems in the academic literature, no pre-set and pre-tested methodology was available to employ in the context of this research. This has allowed for greater freedom in choosing suitable research methods, tools and techniques, and has provided an excellent opportunity to be creative throughout this project.

Given that the aim of the research is the discovery of translation professionals' actual needs with reference to the existing TM systems, the research is geared towards an empirical investigation of needs, based on data that derive from and measure the translation professional's work reality.

The investigation starts with no predefined formal hypothesis. Although research in the area of translation studies or translation technology can be characterised by substantial

knowledge of problems pertaining to existing TM systems, it is not as yet noted for possessing substantial knowledge about hypothetical needs (i.e. explicitly testable hypotheses). In fact, the majority of clear references to translation professionals' needs are use-context specific. For instance, LISA's survey [Lommel, 2004] reveals and briefly analyses the needs of translation companies in reference to TM systems, but the vast majority of respondents are not necessarily the end-users of the systems (instead, they are decision-makers), and in addition, the survey does not give any consideration to non-corporate users, such as freelancers. Other examples of limited references to user needs include the research conducted by Carter-Sigglow [2004] on the translators working with TM systems at the Research Centre Jülich in Germany and the one by Drugan [2004] on translation support tools in the European institutions. On a broader spectrum, more general and diverse needs (that are clear and not mere hints) seem to be fragmented across literature of questionable reliability (e.g. translators' magazines, industry journals, newsletters, etc.).

The lack of pre-existing theory in the form of a hypothesis reinforces the exploratory nature of the current investigation whose goal is to examine the needs of translation professionals in depth and acquire knowledge to form hypotheses that can be tested in future research. Data, therefore, is not used to test a theory, in this context, but to develop a theory. According to the Grounded Theory by Glaser and Strauss [1967], to which the present investigation is close in spirit, a hypothesis at the beginning of a piece of research is inimical to the generation of a genuine theory reflecting reality. Having a hypothesis prejudges the results of the investigation, which severely restricts the generative aspect of the research and consequently the theory [Glaser, 1998: 8-9]. Ideally, research must begin free from speculation and from the literature's claims and be as open as possible to discovery and the emergence of concepts, problems and interpretations from the data [Glaser, 1998:67]. Indeed, this research needed a methodology that could get through and beyond conjecture and preconception to delve into the concerns and visualisations of translation professionals in order to uncover their needs.

The Grounded Theory is a popular research methodology in social research, which could have been considered suitable for the purposes of the current project. It has been offered by its inceptors (Glaser and Strauss) as a systematic way of generating a theory from real data by making use of rigorous methods of data analysis. Its main doctrine is "all is data", and its strength lies in its open-ended, discovery-based orientation which relies on methods that, instead of forcing meaning on the participants, are interested in listening to their genuine meanings, grasping their perspectives and studying their concerns and motivational drivers [Glaser, 1998:32]. The appeal of Grounded Theory derives mainly from the fact that the emergent theory is fully grounded on real-life data, which leaves little room for challenging the theory, if the data is fit. Despite its record of success in discovering the reality in the

context of various social research projects, and although there has been evidence of its effective application in Software Engineering as well [Allan, 2003], Grounded Theory cannot be used as a general research methodology in the context of the present research. It can – and will be – used instead as a data analysis methodology in the analysis of qualitative data (see Chapter 4: Analysis of qualitative data).

Its main weakness as a research methodology stems from its lack of internal and external control of its data sources, which may have an adverse effect on the quality of data. According to Glaser [1998:8-9], data can be collected from anywhere and by any means, ranging from the briefest of comments to the lengthiest interview, and it can vary from factual descriptions to airy ungrounded conceptualisations. Clearly, in Grounded Theory there is no procedural structure in data collection, and no standards as to what constitutes useful, relevant and representative data. Any data used, however collected, appears to be of equal value [Glaser, 1998:43]. This lack of control over the data sources and quality, which undermines the rigour of data collection, carries the danger of founding a theory on unfit data, with the serious possibility that the whole theoretical construction will, at some point, collapse.

In the context of this research, a methodology was needed that included rigorous methods, tools and techniques at every stage of the research. These stages are:

1. **Data collection**: elicitation of information from users of TM systems
2. **Data filtering:** evaluation of information in terms of relevance and usefulness
3. **Data analysis:** systematic identification, through coding, of user needs that lead to the discovery of new application fields for TM systems; sorting data into categories
4. **Data interpretation:** organisation and integration of categories into substantive and formal theories; linking the findings to related literature

The four stages are consecutive, reflecting a linear structured process of investigation. Every stage has its own goals and challenges that need to be considered when selecting suitable methods, tools and techniques. The selection is not determined *a priori*, before defining the goals and the problems of each stage. In this way, the methodology allows for flexibility in considering various suitable methods, tools and techniques from different disciplines depending on specific purposes, and it is built gradually, as every stage of the project progresses.

The reasoning that is followed is inductive, in the sense that investigation starts with specific measures of the likes, dislikes and suggestions of translation professionals regarding TM systems, and as it progresses it begins to detect patterns and regularities, formulates some tentative hypotheses that are explored, and finally ends up by developing a

general theory. Inductive reasoning, by its very nature, is more open-ended and exploratory, which fits ideally with the investigative character of the present research.

2.2 Needs investigation in Software Engineering

In this research, the goal of constructing a theory about the translation professionals' needs is to support the development of improved TM systems. An understanding of the development process of software, and of TM software in particular, is helpful in explaining where needs identification fits in that process and leads to the recognition of strategies followed in real-life software development projects to discover user needs. An insight into TM software engineering also helps to clarify the key concepts of *needs* and *users*, which is necessary before embarking on the investigation.

2.2.1 Software development lifecycles

In a typical software development lifecycle, identifying user needs is part of the *requirements definition* stage. This stage is responsible for defining what services are required from the system. The necessity for user needs investigation arises not only at the early stages of translation systems development, but throughout the software development lifecycle provided that the system is maintained through the release of updated versions.

There are nearly as many software development methodologies as there are companies developing software, and in many cases software is developed without any methodology and even without any specifications. This applies to TM developers too, as each one follows their own system development model. In general, though, all models comprise the following basic stages, as illustrated in Figure 4: *requirements definition*, *design*, *implementation*, *testing* and *operation* [Sommerville, 2004: 64-8]. Once the system enters the market, the software development process normally continues in iterations, so that the system evolves around the changing requirements of its users. Ultimately, the technology usually fuses with other technologies to create a new kind of technology, or it dies after being replaced by a better technology.

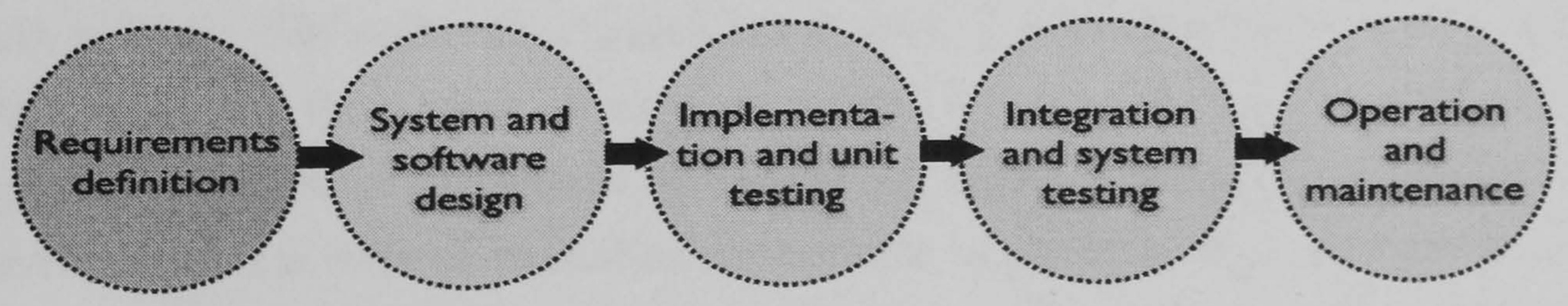


Figure 4: Typical software development lifecycle

According to Wallin and Land [2001], there are three basic types of system development lifecycle model: the *sequential*, *incremental* and *evolutionary* models.

The sequential models were the first to appear and are still largely used by organisations. They build on the assumption that the problem to be solved can be completely understood and described before a solution is designed. A design that satisfies all aspects of the problem can be specified before implementation [Wallin & Land, 2001]. Normally, this approach is applicable when experienced developers work in a well-known problem domain, with well-known technology. Within this model range fall the commonly used *Waterfall* and *"V"* models, but these models seem not to be preferred by developers of commercial TM software. The most probable reason is that translator support technology is a relatively new research area, so that the application domain has not been explored to the extent that it can provide complete user requirements at the beginning of a new TM software project.

The incremental models, with a view to minimising risk, favour the decomposition of a development project into several smaller sub-projects (increments). In these models, as in sequential models, the overall requirements of the final system are known at the start of the development. In incremental models, however, a limited set of requirements is allocated to each increment and with each successive (internal) release more requirements are addressed until the final (external) release satisfies all requirements [Wallin & Land, 2001]. TM developers with many years of experience in developing TM software seem to prefer this software model. Since they already have a wealth of requirements at the beginning of a new TM software project (or a new version), TM developers break them down into sets and allocate each set to a sub-project. Usually, each sub-project deals with a separate module of the software (e.g. terminology manager, alignment tool).

Finally, the evolutionary models are often used by developers of new technologies or developers with little experience in the application domain, when all the requirements cannot be known in advance. The most frequently used evolutionary model is the *Spiral* one, where a prototype is refined to a system in a series of cycles. This is, essentially, the model that most new TM system developers seem to use. All the requirements of the final system are not known when the development starts. Initially the requirements are partly defined, and then refined and extended with each successive release as the requirements picture matures. The main risk encountered by these models relates to the architecture of the first version of the system, which must support the changes introduced in each cycle, otherwise the system will need to be redesigned completely. Admittedly, at the early stages of development it is difficult to foresee all the kinds of problems that will arise so that the developer devises an effective mitigation plan from the start. But if a careful investigation of user needs and clear specification of requirements precede the design phase, there are good chances that many mistakes will be prevented from happening, and overall, the developer

will gain from a more flexibility-conscious approach to the architecture's design and implementation. Evolutionary models are also more suitable for small to medium-sized systems [Somerville, 2004: 69]. Frequently, TM developers abandon evolutionary models in favour of incremental ones, once the system and the developing capacity of developers has matured.

2.2.2 Requirements Engineering

Virtually all system development lifecycle models start with the requirements definition stage, whose complexity has given birth to a sub-discipline within Software Engineering called Requirements Engineering (RE). Requirements Engineering is seen as the bridge between the developer and the user, as its role is to transfer knowledge from the application domain to the development team. While Software Engineering focuses on 'designing the thing right', Requirements Engineering concentrates on 'designing the right thing' [Sutcliffe, 2002], in the sense that it ensures that the requirements reflect accurately the needs of all the stakeholders of the system and that are in such a form that engineers in subsequent development stages can understand them and turn them into a software that is fit for the intended purpose of use. In more abstract terms, Requirements Engineering lays down the 'promise' for the new system (or the new version of a system), whereas Software Engineering is responsible for delivering on that promise.

Requirements are critical to the success of a software project. If the initial requirements for the project are done well, the project has a fair chance to succeed regardless of its size. If the requirements are done poorly and are filled with errors and uncontrolled changes, the project has a great chance of running out of control (if not dropped out).

Requirements are the descriptions of the services to be provided by the system and its operational constraints [Sommerville, 2004:118]. The requirements reflect the needs of the stakeholders for a system that helps solve some problem, such as retrieving a suitable translation example for a given source text from a repository of bilingual aligned texts. From an exploratory study of SE literature the term *requirement* does not seem to be used in the industry in a consistent way. Some use the terms *requirements* and *needs* interchangeably [e.g. Herzwurm et al., 1999; Krogstie, 1999], implying that the two concepts are synonymous, while others make no distinction between *stakeholder requirements* and *system requirements*, and use the term requirements to refer to both [e.g. Ryan, 2000], even though they have distinct differences and are usually represented in two separate documents. According to Sommerville [2004:118], some of the problems that arise during the RE process are a result of failing to make a clear separation between the above concepts. It is important, therefore, to clarify their difference, for the purpose of this thesis as well.

According to Oxford English Dictionary a *need* can be defined either as a *condition requiring relief* (which presupposes the existence of a problem) or as *anything that is necessary but lacking*. Needs are often confused with wants (= desires), which are more discretionary. The present research is mainly interested in actual needs, whether those are expressed explicitly, or are hidden in various demands or desires. The users' desires are also important for this research as, although they may not stem from immediate obvious problems, they can be the needs of the future, especially if they derive from users who, being savvier with IT, have become more sophisticated over time.

Needs are not solutions to problems. They are usually the problems themselves. The user need is the *basis* for a user requirement, which in turn corresponds to a system requirement. Needs are expressed in layman's terms and because of the inherent problems in natural language interpretation, they need some processing and analysis to become user requirements. The greater part of RE literature appears to disregard or gives very little detail of the process of user needs identification and analysis as a separate step within user requirements engineering. Nevertheless, the analysis of needs is of critical importance in an RE process, as correct user requirements can only derive from correct identification of user needs.

In the context of this research, the identification of user needs is an essential step before forming user requirements for TM systems, as in order to understand the visualisations of users about their ideal TM system through the perceived solutions for their immediate problems, it is important to analyse the situation that can reveal a bigger problem, for which the problem that users see may only be a symptom. The analysis of a situation is, thus, a way of ensuring that the correct need is identified through the understanding of the reasons behind given perceived solutions.

Another reason for identifying needs before documenting user requirements is that user requirements, as perceived solutions, are usually tied to the technological capabilities of existing TM systems and therefore are time-bound and are highly likely to change within a brief period of time. Needs, on the other hand, because they are independent of specific technological solutions, withstand longer and change at the same rate at which the translator's profession transforms in the socio-techno-economic continuum. Therefore, the identified needs can be used safely over and over again as the basis for continuously renewed user requirements for longer periods of time.

While user requirements are independent of the technological, financial and organisational constraints of the TM developer, the system requirements describe the solutions to the users' requirements that are feasible within the developer's capacity. System requirements are more technical in nature and more detailed in the description of the system's required services [Somerville, 2004:118-9]. The level of detail is very important in the system

requirements specification. System requirements must be precise and detailed enough to describe a desired and feasible solution but not too detailed, as this might restrict the imagination of the system designer, and cause creative opportunities to be missed, at the subsequent stage of software design.

Table 4 summarises the definitions for key concepts in RE and provides an example for each concept as it appears in a typical set of TM software documentation. It is interesting to notice here the evolution path that a user need normally (or ideally) follows in order to become a product feature.

	USER NEEDS	**PRODUCT FEATURES**
Definition	Statements by the users describing a problem that needs to be solved or anything that is necessary but lacking	Characteristics of the final software
Example	"I need to be able to use my TM database in a different TM system"	TMX (Level 2) compliance
Term used in RE	**User requirements**	**System requirements**
Definition	Brief statements about perceived solution(s) by the user (with no consideration to cost or technology limitations)	Brief, concise statements about the feasible solution proposed by the engineer
Level of detail	High-level abstract description	More detailed description
Example	"The user requires the ability to export the TM database in a standard format, so that the database can be imported and used effectively by other TM systems"	"The system shall comply with the TMX standard"

Table 4: Important definitions in Requirements Engineering

Very few companies have an explicitly-defined and standardised RE process [Kotonya & Sommerville, 1998: 9]. Moreover, different companies tackle requirements engineering in radically different ways. For example, Kovitz [2003] observes that there are significant

differences in requirements engineering between phased (plan-driven) and agile development. Whereas phased development does requirements engineering in an early phase that precedes the majority of coding, agile development runs the process throughout the project. In addition, phased development usually embodies the requirements in a written document, while agile development does not.

TM developers also have different ways of dealing with requirements. The following two case studies[31] demonstrate the differences in RE practices between a large company developing TM systems, which follows a phased development approach, and a small development team which develops an open-source TM system by following an agile approach.

Case study A: RE for a leading commercial TM system

- In our company we follow an incremental development process. This has proven to be the most efficient process to ensure that we ship product that has the required levels of stability, quality and robustness when it is released to the market. We find it especially important to make quality assurance an integral part of the development process and have tight links between core engineering and stakeholders from quality assurance and documentation functions to ensure that the product we ship is not only "coded" right but has also the right quality and comprehensive documentation.
- We have a formal documented procedure within the company for user and system requirements specification. We use an approach that is revolves around MRD (Market Requirements Document), PRD (Product Requirements Document) and ERD (Engineering Requirements Document). The product marketing function is responsible for the commercial requirements driving a product inception/enhancement via the MRD, which then feeds into product management for the more technical PRD (describing concretely what features are required to fulfil the user benefit requirements from the MRD), which in turn feeds into engineers writing ERDs for individual end user features.
- Specifying use cases is an important step throughout the PRD, ERD and QA process. It is important to understand how the user will actually use a feature, what the desired results are and what potential exceptions might occur.
- The people who are involved in user requirements identification come from: Product Marketing, Product Management, Sales, Support/Professional Services, and Business Consulting. Gathering user requirements is an ongoing effort that never stops. The primary

[31] The information is reproduced from the completed questionnaires received from the Product Manager of Company A and the Lead Developer of Development Team B in April 2006. The names of the developing company/team, of the TM software products and of informants are not disclosed for confidentiality reasons and to avoid the possibility of generating positive or negative publicity for the specific products.

methods that we use to identify user needs are: feedback/complaints from users, user visits, and discussions that happen with users during shows/conferences. Occasionally, we conduct user surveys, we run focus group sessions and/or gather information from existing marketing data.

- After user requirements have been discovered, they are fully evaluated and, if feasible and reasonable, feed into both the Market and Product Requirements Document. The prioritisation of requirements is carried out by Product management, in cooperation with key stakeholders from development management, product marketing, sales and customers in focus groups.
- We cope with changing user requirements by closely involving key customers and focus groups in shaping a feature, getting feedback and changing it as required, and by working closely with architects and key engineers from the development function to ensure the infrastructure for a feature is sound and future proof, rather than merely implementing user feedback tactically.

Case study B: RE for an open-source TM system

- We do not have a strict software development process. We're following "Release Early, Release Often" development model, which is close to an evolutionary process. The reason for choosing one is that we're an open source project where developers are free to join and go. The main benefits are that we get lots of feedback and we get it early, and that this allows us to implement easy improvements more or less ASAP, while being able to plan long-term improvements in advance.
- We have a formal documented procedure within the company for user and system requirements specification, which is a public RFE (request for enhancement) report that anyone can fill in.
- User requirements are identified most of the time by power users who are part of the community, but as well as by other users. Developers are responsible for evaluating the public RFEs. Gathering user requirements is ongoing. We elicit user requirements primary from the feedback/complaints we get (through RFE or e-mails). Frequently we conduct focus group sessions, and occasionally we interview users individually or gather information from the conferences/shows we attend. Rarely, we conduct user surveys.
- After the discovery of user requirements, we follow the following steps: a) evaluation, b) assignment to a particular developer, c) implementation, d) evaluation of the implementation against the user's request. We do not use technical specifications. The prioritisation of system requirements is performed by the lead developer and other

developers and our criteria are mostly subjective: "do I like it?" and level of difficulty.

- In an effort to cope with changing user requirements, we try to discuss the requirements thoroughly before implementing them, but if the evaluation of the implementation against the requesting user shows the requirement is void, we lose.

The two case studies highlight two different approaches to RE based on the resources that each developer possesses. The first approach is more systematic and several people (including users) are involved in the different stages of the RE process. Requirements are collected through various methods and are negotiated in consultation with the users. The second approach, on the other hand, is less systematic but flexible to cope with the peculiarities of an open-source project. Minimum effort is put into the RE process by concentrating on few but effective methods of collecting requirements, and by evaluating requirements in a swift process with no further consultation with the users requesting the enhancements. The advantage of this rapid agile approach to RE is the prompt implementation of requested features. The former approach, on the other hand, concentrates on the features that would be desirable to the majority of stakeholders, and not all the features requested, which is why the developer devotes more effort to the negotiation of requirements.

In general, most software developers try to follow a systematic route to RE and organise RE activities in groups. Kotonya & Sommerville [1998:32-3] distinguish four main activity groups in a typical RE process as illustrated in Figure 5, but point out that these sub-processes have no distinct boundaries, as they usually run as an iterative process where each activity provides feedback to the others. The four sub-processes are:

- **Requirements elicitation**

  This is a human-centred group of activities during which needs are discovered and stakeholder requirements are collected through interaction with the stakeholders of the system. Requirements are also collected from existing system documentation, from market studies and related standards and regulations. Domain information is also collected during this stage from stakeholders or relevant literature. This process is also referred to as *requirements acquisition* or *requirements discovery*.

- **Requirements analysis**

  This group of activities includes *requirements classification*, *organisation*, *prioritisation* and *negotiation*. The unstructured collection of requirements from

the previous stage is broken down into groups of related requirements and requirements are organised in coherent clusters (e.g. by type of stakeholder, by user group or by system module). Inevitably, where multiple stakeholders are involved, requirements will conflict. Therefore, at this stage, the requirements are prioritised in consultation with the stakeholders, and any requirement conflicts are resolved through negotiation.

- **Requirements documentation**

  The agreed and prioritised stakeholder requirements are documented at an appropriate level of detail in a document that represents stakeholder requirements. Based on this document, the engineers of the system might need to add more technical details, so they either create a separate document for system requirements, or add their input straight to the final *requirements specification document*. Formal or informal requirements documents may be produced at this stage, and can be in the form of documents in natural language or/and diagrams.

- **Requirements validation**

  The final requirements are checked carefully for consistency and completeness. The process is intended to detect problems in the specification document before it is used as a basis for the system development. In case of a problem found, some of the previous activities may have to be repeated.

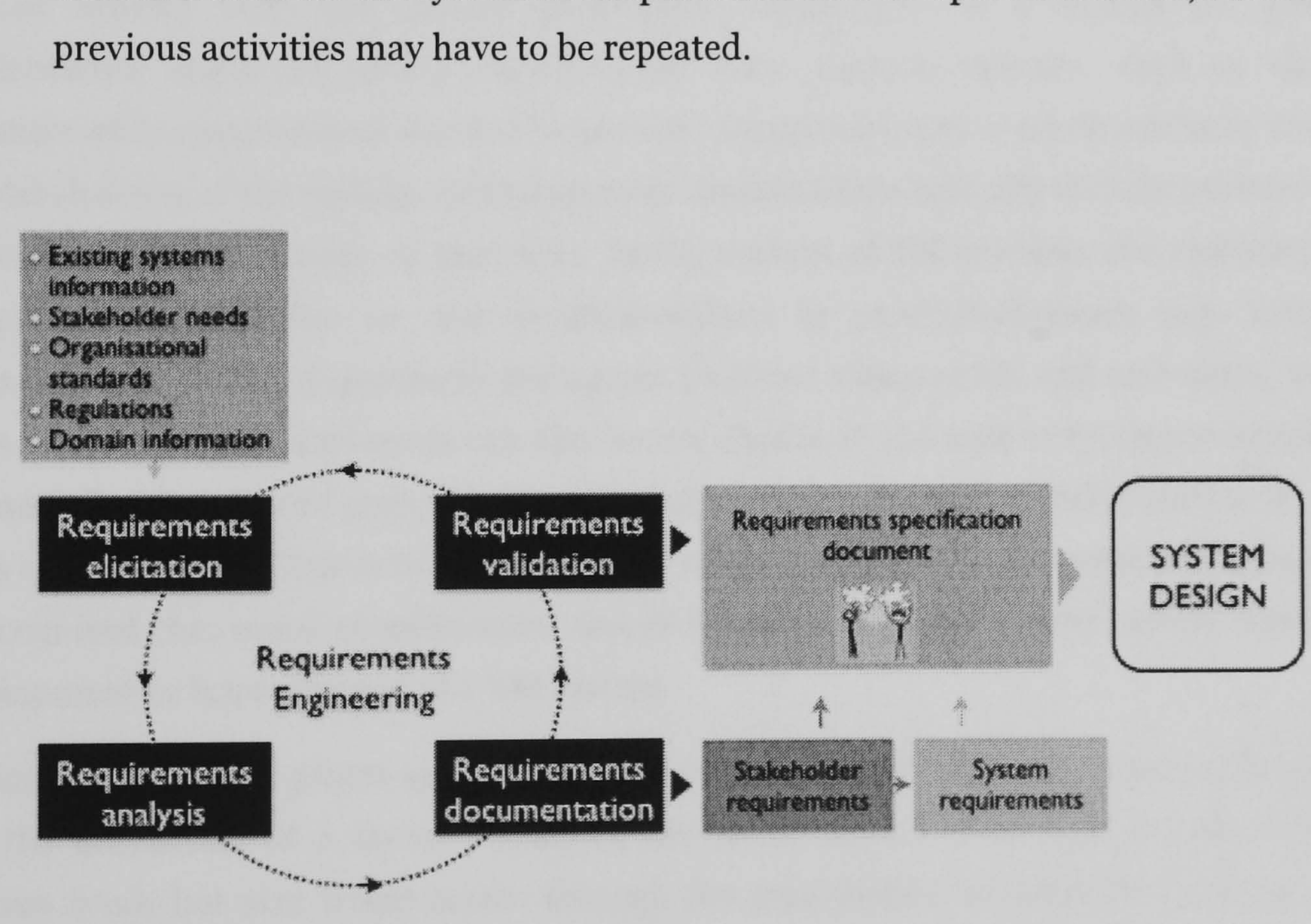


Figure 5: Typical Requirements Engineering process

The goal of the requirements engineering process is to create and maintain a *requirements specification document*. This document plays a central role within the TM system

development process, as it feeds directly into system design and is regarded as the main instrument that guides all the other project actions. Jackson [1995] clarifies the relationship between the requirements and the design stage by pointing out that requirements are mostly concerned with the problem to be solved, while design is concerned with the solution to the problem [Kotonya & Sommerville, 1998: 10]. In other words, and in the case of TM systems, the requirements set out what the stakeholders – including users – require from a TM system, whereas design describes the path to be followed to meet these requirements.

In the context of this research, the typical RE process validates the line of investigation (see section 2.1) followed in this project. The four stages of *data collection*, *filtering & organisation*, *analysis* and *formalisation* correspond to the RE sub-processes *requirements elicitation*, *analysis* and *documentation*, to the extent that the current project could be characterised as an attempt to perform requirements engineering for an improved type of TM system. However, despite the common goal of improving a TM system through new requirements, there are crucial differences between the present project and a typical RE process for a TM system.

First of all, this project is interested in requirements that derive exclusively from *user* needs; in other words, the focus is on *interactor viewpoints* represented by the people who interact directly with the system [Somerville, 2004:150]. In a typical RE process, requirements engineers gather requirements from various sources, such as domain literature and organisational standards (*domain viewpoints*) and most importantly from all the stakeholders of the system, including users. Stakeholders typically include customers on one side and the developer on the other. In the context of TM systems, the customers are categorised into clients, i.e. the decision-makers or product-choosers (e.g. company managers, translation department managers) (*indirect viewpoints*) and end-users, i.e. the users of the software. End-users can also be the clients, in the case of freelance translators for example. The present project is interested in the later category of stakeholders: the end-users of TM systems. This is mainly due to the scope limitations of this research, although it is recognised that equal consideration should be given to all customers' needs during the development or improvement of a TM system.

Secondly, this project places special emphasis on the in-depth analysis of user information and the generation of a theory regarding the users' actual needs that reveals not only obvious needs but also latent needs, through the examination of causality between needs and the user's situation. This in-depth analysis is a step that is lacking in RE projects due to previously mentioned constraints related to cost and time limitations (see p. 52). However, it is important that this investigation goes beyond the superficial identification of user needs, as performed in a standard RE process, to unlock hidden messages that could foretell the future of TM systems.

Thirdly, the present investigation ends in a theory around identified needs that is grounded in user data and can be the basis for formal user requirements. To reach the final stage of producing a TM system design, developers need to translate the identified user needs into requirements, to contrast them with requirements from other sources, to negotiate them in case of conflicting requirements, and to validate them against system requirements. All these subsequent stages necessary in a RE process fall beyond the scope of this research. If the scope of research could be extended, such research could develop further with a potential collaboration with a TM developer who would provide information on other sources of requirements (e.g. marketing data, requirements for previous versions of the system) and would offer the developer's viewpoint during the negotiation and validation of requirements. However, this possibility has been rejected for the present research project on the grounds of intellectual property rights protection[32]. Consequently, the current research ends when its aim has been achieved, i.e. once it succeeds in presenting the needs which reveal areas of improvement in TM systems, thus offering critical information in the application domain of TM systems that TM developers are lacking.

Based on the above, the investigation route followed in this research has many differences from an RE process for TM systems. Nevertheless, the present investigation could be considered as *part* of an RE process for TM systems, since most of its steps overlap with RE sub-processes.

One of the first steps in RE, which is also relevant for the preparation of this investigation, is the identification of users, who are the exclusive informants at the data collection stage. During this identification, it is important to notice any distinction between different user groups and any special interests that the user groups may have. The investigation will then be able to focus on those interests when eliciting information from different user groups. It is also essential to determine which group is more important than the others, in order to place greater emphasis on input deriving from this particular user group.

2.3 Definition of TM users

It is widely observed that translation professionals are normally the only users of TM systems. Translation professionals include, apart from known professionals such as translators, terminologists, proofreaders and reviewers, every professional who is involved in the production of translations (such as project managers, graphic designers/engineers,

[32] The outcome of the present investigation has been evaluated by Imperial Innovations Limited (consulting company, operated by Imperial College London, offering services related to the commercialisation of research produced at Imperial College London) as holding significant commercial value, therefore the disclosure of data to a third party could be exploited financially. Further to College's recommendations, it was opted not to give access to data deriving from the present research to any TM developing company before the research is published, as this could be seen as favouring a specific developer over the others.

DTP specialists). This typology of TM users is without difficulty confirmed by browsing the websites of TM developing companies and noticing who their systems are targeted at. It appears clear that from the above user groups, translators are the most significant user group, possibly because TM systems were originally developed especially for them. The investigation is, consequently, more interested in the needs of translators rather than those of other translation professionals, without this meaning that other translation professionals are excluded from the investigation or their needs are overlooked. In fact, their needs are studied too, and compared with the needs of other user groups; however, the information acquired from these user groups is limited in quantity, compared to information acquired by translators.

The present investigation studies all three types of TM end-users:

a. **potential** users: people who are involved in the production of translations or in any aspect of a translation workflow, but do not currently use a TM system

b. **emerging** users: people who have not used a TM system in the past, but are known to have the intention of using a TM system in the future

c. **actual** users: people who are using a TM system at present or have used one in the past

The main reason for considering potential and emerging users in addition to the actual ones is that translation professionals have needs independently from a TM system, so the needs should not be related to whether they have used a TM system before or not. Obviously, for a person who has been using a TM system it will be easier to express her needs if she thinks of her known TM system as a model of reference. But it is highly likely that this person's needs will be prejudiced by the limitations of the existing technology. On the other hand, a person with no preconceived notion of a particular translator-support solution will be able to "think outside the box" and generate more daring ideas about future solutions.

Finally, the investigation pays particular attention to lead (or "power") users as they are a better indication of future needs [Von Hippel, 1988]. Lead users are actual users who generally have lengthy experience in using TM systems (typically use – or have used - more than two systems) or use the system to its full potential for various tasks. According to Von Hippel's study [1988], lead users are more valuable informants than the other users because they currently experience needs still unknown to the public. Because they are able to express needs months before the rest of the marketplace encounters them, they are regarded as a key source of innovation.

2.4 Methods of data acquisition

Following the identification of informants, a critical step in the preparation of this investigation is the review of available methods for sourcing data and the selection of a method that will ensure that the acquired data is fit for the aim of this research.

2.4.1 Sources of data

Two directions could be followed in order to address the question of sourcing data. The data can be any information deriving from users themselves, or from any other source that possesses relevant information, related to what users (might) need, want or expect from a TM system. Thus, one way to investigate translators' needs would be to collect data from secondary sources, such as literature, which involves the identification of relevant data, its evaluation and analysis, in order to generate the theory that answers the chief question of what translators need in reference to TM systems. Secondary data can be acquired from:

- Academic literature (books, conference papers, academic journals)
- Practitioner journals, magazines, newsletters
- TM users fora
- Translators blogs, TM users blogs
- TM developers' documentation/websites (from which statements of needs can be extracted)

In particular, valuable information on user needs can be gathered from TM system evaluations carried out by users, and system independent reviews that are either comparative (comparison of features for two or more systems) or standalone. However, these sources prove inadequate in view of the following problems:

- In system evaluations by users, there are descriptions of what is good and what is bad in existing systems, but hardly any new ideas are ever brought out.
- In system independent reviews, the authors hardly ever propose alternative solutions, and even when they do, it is hard to trace the rationale behind the proposed solution, and hence, it is difficult to see the reasons why one particular solution is more preferable than others. In addition, they rarely ever provide reasons why some features are not desired (which would possibly reveal a need). By nature, reviews published in magazines or journals lack depth in their analysis because of length constraints, so the data provided is limited in detail.

Generally, relying on secondary sources of data is problematic because of the following issues:

- Quality of sources: the majority of sources containing relevant information (apart perhaps from books, reviewed journals and conference contributions) cannot be considered authoritative in the sense that it is questionable whether the relevant information is true and accurate.
- Representativeness of data: theories on translators' needs may draw from data collected from a sample unit that is not representative of all TM users.
- Time sensitivity of data: needs change over time, as translators become more competent in the use of computers and as new improved systems become available to them. Old sources of data are likely to carry obsolete information about needs that have already been covered by the majority of TM systems. In addition, because needs reflect the translators' work reality in a certain period of time, the information that one needs to study this time-confined reality must be synchronous.
- Control over sources: there is little control over the measurement and the results presented may be tainted by data of poor quality; it is also difficult to trace back the empirical evidence that was used to generate the theories that the authors present.
- Methodology: the literature is full of claims and theories that are not generated through a systematic approach.
- Subjectivity: the theories presented in the literature are usually of a subjective nature, or are biased in favour of (or against) specific TM systems. Especially in the case of system reviews and evaluations, the literature commonly reflects the opinion of a single person.

Due to the above inefficiencies and deficiencies of secondary sources, the present investigation is geared towards a more controlled process of data collection that does not have to consider any particular data as sanctified, objective or valid, and, at the same time, one that places research closer to the real world. Instead of relying on *in vitro* paradigms of the translators' work reality, the investigation has set out to collect primary data, directly from the informants, in order to be able to base the theory on *in vivo* measurements.

Acquiring information directly from users, rather than mining information out of documents, is not intended to diminish the importance of documentation or literature. Information acquired through communication with the users will be interpreted with the help of existing literature, and literature will be contrasted with the theory that will emerge out of the data. No theory stands alone; it must be integrated into the yet bigger picture of the area of Translation Technology. For this to happen, the scholarly incorporation of literature needs to relate the theory to the literature, so that literature will legitimate even more the grounded theory as a contribution to this area.

2.4.2 Available elicitation methods

The immediate question that occurs as soon as users are established as the exclusive source of data is *how* to elicit knowledge from them. Requirements Engineering offers a variety of methods for capturing user information, whose fitness needs to be evaluated according to criteria relevant to this research. The primary criteria informing method selection in the context of this project are:

- Research goals: the method must allow a large-scale collection of data (textual and numerical) that should expose or cue the needs of translation professionals in reference to TM systems
- Cost: it must bear minimal or no financial cost due to budget restrictions
- Rigour: it must be rigorous and must represent a systematic data acquisition programme
- Appropriateness to domain: it must have relevance to the people of the domain

Some of the most widely used elicitation methods in Software Engineering and other disciplines are discussed below, and their appropriateness for this research is assessed.

2.4.2.1 Observation

Observation is a passive data acquisition method commonly used by requirements engineers in order to uncover and understand a) implicit needs [Kotonya and Somerville, 1997:69] and b) needs that pertain to specific social or organisational contexts [Somerville, 2004: 157]. The analyst observes users in their day-to-day work, while using the TM software or while discussing about it with their colleagues, and keeps comprehensive notes. Somerville points out that the value of observation lies in the fact that people usually find it very difficult to articulate details of their daily work because it is second nature to them. They understand their own work but may not understand its relationship to other work in the organisation. Therefore, social and organisational factors that affect a system's use but are not obvious to individuals may only become clear when noticed by an unbiased observer. Kotonya and Somerville argue that observation is a better way of understanding tasks that users are performing in an intuitive way than direct questioning. This is because when tasks become routine, users do not have to think about them consciously and often they may not even realise what they are doing.

In the context of this research, passive observation could not be used as a data acquisition method for a number of reasons. Firstly, observation is reported to carry the risk of capturing large amounts of irrelevant data [Maiden and Rugg, 1996:184]. Secondly, effective observation of users can only happen in their workplace, but, despite the fact that gaining access to work places can be problematic, the use of TM systems would have to be

studied by taking into account organisational factors. However, the aim of the present investigation is to reveal user needs independently of a specific organisational context. Finally, although observation could provide information on the application domain of TM systems (e.g. how particular tasks are performed in a TM system, what problems are encountered), it is not certain whether substantial information hinting at possible needs would be gathered.

2.4.2.2 Think Aloud Protocols

Protocol analysis is another method combining observation with verbal communication, frequently used in Requirements Engineering, especially for building expert systems [Ramesh and Jarke, 1999:7], and even more frequently in Translation Studies, as a Cognitive Task Analysis (CTA) tool, for investigating translation as a cognitive process or for identifying translator strategies [e.g. Lörscher, 1996]. With this method, the thought processes of the user are captured by recording the subject speaking her thoughts out loud. For instance, the user can undertake particular tasks with a TM system and speak out loud during these, explaining and commenting on every step of the process. The advantage of this method is that it can provide large amounts of data about specific tasks, while its disadvantage concerns the fact that transcribing and analysing protocols can be tedious and time-consuming [Maiden, 1996:184]. However, the main reasons why protocol analysis has not been used as a data acquisition method in this research is to avoid being tied to a limited number (six maximum[33]) of TM systems for the study – based on the data for which it would not be possible to make generalisations for all TM systems – and because this research investigates user needs with reference to a complete TM solution and not to a particular TM task. Inspecting user-software interaction for specific TM processes would significantly limit the chances of identifying unique areas of improvement for TM technology.

2.4.2.3 Scenarios (user stories)

Creating scenarios is yet another method for eliciting requirements from users. Its success as an elicitation method lies in the fact that people find it easier to relate to real-life examples than to abstract descriptions. They can understand and critique a scenario of how they might interact with a system [Somerville, 2004:153]. Scenarios can be particularly useful for adding detail to an outline requirements description created by a requirements engineer. They are descriptions of example interaction sessions, covering one or more possible interactions between an end-user and the system. Applying a scenario involves the requirements engineer and the system end-user working through the scenario together,

[33] Imperial College London, in the period 2004-07, has owned licences for six TM systems (most of which are standard editions) that could have been used for this research.

with the engineer taking notes of the user's comments, problems and suggestions. The end-user simulates the use of the system following the scenario and points out areas where the scenario is incorrect, simplistic or variable. The requirements engineer may ask questions at various points about current user actions, how tasks are carried out, who is involved in tasks, and what would happen if some alternative approach was taken.

Using scenarios to describe requirements is an integral part of agile development methods, such as extreme programming (XP) [see Beck and Andes, 2004]. In XP, scenarios are called user stories instead, and they are usually written jointly by programmers and on-site users on index cards. Each index card provides a concise description of the system and the desired functionality for the system under development. Each card also represents a unit of user-visible functionality, or in other words a user requirement [Beck and Andres, 2004:44-5]. An XP-style elicitation method would not be appropriate in the context of the present research mainly because it is not possible to engage TM programmers in this process (see section 2.2.2, p. 64, for the reasons). In addition, any scenario-based method presupposes the generation of an outline description of the system and its functionality, corresponding either to an existing version of a system or to a system that is within the development capacity of the developer, that the end-users can work on. In this research, the investigation of user needs that will lead to user requirements needs to remain free from any particular existing TM solution, in order to be able to reveal deeper unmet needs and more original ideas about future features.

2.4.2.4 Prototyping

People possess tacit knowledge about their needs that is not accessible to conscious introspection, and that is often missed when using typical elicitation methods involving verbal communication (e.g. interviews, focus group discussions). The use of a software prototype allows a simulation of the required system and its interaction with the user, and provides effective cues for recalling and recognising tacit knowledge [Maiden and Rugg, 1996:188]. For example, prototypes can often remind the users of low-level functions that are often otherwise overlooked, and can expose omissions, errors and inconsistencies in the software under development. This way, user needs are easier to discover. The present research has opted not to have a prototype built and used for the purpose of eliciting tacit knowledge mainly because of the high cost involved and because of problems present in engaging external resources for its development. Instead, this investigation has opted for an alternative technique for eliciting tacit knowledge, through the elaboration of questions that help the introspection of translation professionals with regard to their work practices (see Chapter 3).

2.4.2.5 *Focus group discussion (brainstorming)*

Focus group discussions, or "brainstorming" sessions as they are commonly referred to in marketing, represent another method for eliciting information from the users of a system, with a view to capturing new ideas about future desired features. Focus group discussions can be conducted by the software developer themselves or by a market research company assigned to discover new trends and ideas for the software system under development. The session normally consists of a semi-structured exchange of viewpoints and ideas among a small group of users with a clear agenda. Some TM developers often carry out such sessions, during which users comment on the existing TM system developed by the company or simply visualise new solutions independent of the existing product. Focus groups are the mainstay of consumer research and the method is often regarded as one of the most important qualitative data collection methods used in market oriented research. The strength of this method is that it allows a space for the users to be creative and express their ideas, and also to question assumptions that might otherwise have constrained the system's functionality. One of the major concerns for conducting focus group discussions is that participants may not be intuitive enough to contribute to new concept generation, and there is always the risk of collecting unwanted data.

The primary reason for not choosing a focus group discussion as the data elicitation method in this investigation relates to the difficulty in engaging a large number of people to this research and most importantly in organising a series of face-to-face focus group discussions in different locations. The alternative could be to stage the focus group discussions on the web (communication through e-mails, or through a dedicated forum) which would overcome the problem of approaching large numbers of users and that of the physical location of participants. However, this method presents a different set of issues. Firstly, it would be very difficult to maintain the interest of participants during the whole period of a virtual focus group discussion, which means that critical questions could be left unanswered for a long time and the moderator would have to probe answers continuously and until a satisfactory level of information was reached. Given that a significant number of users are needed to provide information, a series of no fewer than ten focus group discussions (8-10 participants) would be needed, which would render designing and coordinating the communication between so many people a very daunting task. As a consequence of all the risks involved in conducting several virtual focus group discussions, it would be difficult to estimate the duration of the task of data elicitation at the beginning of the project and it is highly likely that the data elicitation would delay the entire project.

2.4.2.6 *Interviewing*

In-depth interviews are one of the main methods of qualitative data collection in market research as well as in Requirements Engineering. The interviewer spends time with the

users of a system in one-on-one interviews finding out about their particular circumstances and their individual opinions about the system. Interviews can be formal or informal, and depending on the information they aim to elicit they can be:

- Structured: the interviewer asks a predefined set of questions, so that all respondents cover the same issues. The order of questioning is fixed and wording is usually specific: there is little scope for probing or deviating from the specified agenda.

- Semi-structured: this type is more flexible than the structured interview. Although the interviewer in this technique will have some established general topics for investigation, this method allows for the exploration of emergent themes and ideas rather than relying only on concepts and questions defined in advance of the interview. The interviewer would usually use a standardised interview schedule with set questions which will be asked of all respondents. The questions tend to be asked in a similar order and format to make a form of comparison between answers possible. However, there is also scope for pursuing and probing for novel, relevant information, through additional questions often noted as prompts on the schedule. The interviewer frequently has to formulate impromptu questions in order to follow up leads that emerge during the interview.

- Unstructured: interviewees have the freedom to explain their work routines and how they use TM systems in their own way, although there may be some gentle guidance offered by the interviewer in order to keep the narrative going. The method has some overlap with the semi-structured interview, in that the interviewer may have a very simple schedule, but in the unstructured interviews that may not be strictly followed. In the unstructured interview the interviewee is treated as an active subject, and not merely a reporter of facts or experiences. This method is concerned with finding what the user wants from a TM system through a detailed discussion of the interviewee's professional needs.

In interviews, different techniques are used to elicit the desired information. For example, the interviewees are frequently asked to picture themselves working with a TM package, and to state the problems they encounter while performing specific tasks. They are then asked to offer suggestions as to how these problems could be overcome.

Generally, interviews are regarded as an effective method for gaining insight into the user's perspective and can provide a wealth of information on the nature of user-TM system interaction, as well as the perceived areas of potential benefits. Nevertheless, interviews are considered less effective for uncovering tacit knowledge (i.e. knowledge that users think is too trivial to mention but could be important for the developer's understanding). Other

issues with interviews include poor recall on the part of the respondent and communication of incomplete or incorrect knowledge.

In the context of the present research, interviews are not a cost-effective means for collecting the necessary qualitative data, due to the limited number of users that could be practically reached. Even if the investigation would opt for telephone interviews, or even on-line interviews, the number of interviews would not surpass 50-60. Given that this research is interested in supporting the design of commercial systems, and not of an internal system with a limited set of users, not only is it important that the data to be collected is of good quality but its quantity also needs to be substantial enough to ensure rich and diverse information that will help generate a more complete picture of what translators need from TM systems. An additional weakness of interviewing is that it assumes that the user has conscious, accurate access to all relevant knowledge. However, this is not often the case. Much knowledge is not accessible to conscious introspection, and of the knowledge that is, much may be missed.

2.4.2.7 Survey research

Investigating users' needs is primarily a qualitative research task and as such it requires a method that can collect qualitative information from as many users as possible, and that is fit for and relevant to the domain of investigation. From the pool of available methods offered by the disciplines of Requirements Engineering, Translation Technology, Social Research and Marketing Research (many of the most common of which were discussed in the previous section), a survey has been selected as the most suitable method of data collection.

A survey is a systematic method for gathering information from a sample of entities for the purposes of constructing quantitative and qualitative descriptors of the attributes of the larger population of which the entities are members [Groves et al., 2004: 2]. In social sciences, it is one of the most commonly used methods to understand the way societies work or to gain insight into a social problem [Groves et al, 2004: 4]. Market research also uses surveys to seek knowledge about reactions of "real" people to existing and planned products or services, and to produce information that is useful for commercial decision makers.

Surveys are conducted only occasionally by requirements engineers in the framework of collecting requirements for new systems (see Case Study A). They are usually avoided due to incurring costs in terms of resources and time needed to design them, carry them out and analyse their results. It is also problematic as a method because if it is branded, people might not be honest, and people who use the particular software under research are likely

to be the only ones to reply. There is also the danger of competitors giving wrong answers with a view to undermining the survey.

In Translation Technology, surveys are commonly used to explore the use of translator-support systems by their users [see Zielinski and Safar, 2005, regarding terminology management tools, and Wheatley, 2003, regarding TM systems] or to investigate attitudes towards TM systems [see Dillon and Fraser, 2007]. LISA has also used surveys [LISA, 2004; 2002] not only to probe aspects of TM system use, but also to elicit requirements for future systems – something that the present research is focused on. Additional details of well-known surveys on the use of TM systems are given in Table 5.

Surveys on the use of TM systems			
Carried out by	*Year*	*Sample unit (respondents – coverage)*	*Focus*
LISA/OSCAR (www.lisa.org)	2002	134, world-wide	Corporate use/usage rate of TM systems, TM implementation from a business perspective, context of TM use, pervasiveness of open standards adoption, future trends.
ITI, BDÜ and various other partners in the context of the eCoLoRe project (Wheatley, 2003)	2003	462, UK	TM use rate, context of TM use, training requirements and issues, attitudes towards TM adoption.
Loughborough University (Fulford and Granell-Zafra, 2004)	2004	390, UK	Uptake of online tools and web-based language resources by freelance translators, implications for translator training, professional development, and research
University of Westminster (Dillon et Fraser, 2007)	2004	59, UK	Translators' perceptions of TM, translators' attitudes towards TM adoption, with a particular emphasis on younger and/or less experienced translators.
LISA/OSCAR (www.lisa.org)	2004	274, world-wide	Corporate use/usage rate of TM systems, TM implementation from a business perspective, context of TM use, pervasiveness of open standards adoption, future trends.
Imperial College London (Lagoudaki, 2006)	2006	874, world-wide	Users' perceptions of TM, context of TM use, practices around TM use, TM use/usage rate, evaluation of TM systems, user preferences, user requirements, future trends.

Table 5: Surveys on the use of TM systems

Survey research presents certain advantages that satisfy the project's criteria for method selection (see section 2.3.2).

- The survey is an efficient way of collecting information from a large number of respondents. Very large samples are possible. It allows the collection of both numerical and textual data, which is necessary for the present qualitative research.
- Surveys are flexible in the sense that a wide range of information can be collected. They can be used to study the attitudes of translation professionals, their beliefs, their work practices and any other factor that influences the professional's needs in reference to TM systems.
- They can provide access to variations in larger populations, allowing the distinction and separate study of sub-groups within the surveyed population.
- Because they are standardized, they are relatively free from several types of error.
- They are relatively easy to administer.
- There is an economy in data collection due to the focus provided by standardized questions. Only questions of interest to the researcher are asked, recorded, codified, and analyzed. Time and money is not spent on tangential questions.

A single method of data collection (i.e. the survey) is preferred over the combination of two or more methods in the framework of the current project. Mixing methods can create several problems to the investigation, the most serious of which being the difficulty of effectively comparing information that is collected by different means. In fact, one method seems to be enough for the present research project, as the aim is not to discover all the needs of translation professionals or, for that matter, to achieve user requirements completeness – as a developer should aim for – but instead, this research is interested in identifying needs that uncover areas of improvement for the existing TM systems. The current research is not a TM development project itself, but an investigation that will provide input to any TM development project.

2.4.2.7.1 Survey mode

There are two main methods (modes) for conducting survey research: by distributing a self-administered questionnaire or by carrying out interviews. More specifically, a survey can be conducted by mail, by e-mail, on the web, by telephone or by face-to-face interviews with people, individually or in groups. Research in survey methodology has shown that the mechanics of the data collection procedure can affect survey results [Dillman, 2007: 217; Ornstein, 1998: 67], and, for that reason, there is considerable research comparing the different survey modes, so as to assist those who do not know which mode would best suit their project [see Dillman, 2007: 217-32; Groves et al, 2004: 145-65; Couper, 2000;

Ornstein, 1998: 67-8]. The table below (Figure 6) summarises the main differences between all survey modes.

Issue	*Self-administered Questionnaire*			*Interviews*		
				Face-to-face		*Telephone*
	By mail	By e-mail	Web-based	Individual	Group	
Can the respondent retain her anonimity?	No	Yes	Yes	No	No	No
Is the method flexible?	No	No	No	Yes	Yes	Yes
Are open-ended questions feasible?	Yes	Yes	Yes	Yes	Yes	Yes
Can the quality of response be controlled?	No	No	No	Yes	Yes	Yes
Are high response rates achievable?	No	Yes	Yes	No	No	?
Is it low cost?	?	Yes	Yes	No	No	?
Is it eco-friendly?	No	Yes	Yes	No	No	?
Are facilities needs low?	Yes	Yes	Yes	No	No	Yes
Does it give access to dispersed samples?	Yes	Yes	Yes	No	No	Yes
Does the respondent have unlimited time to formulate answers?	Yes	Yes	Yes	No	No	No
Is there personal contact?	No	No	No	Yes	Yes	Yes
Is a long survey feasible?	Yes	No	No	Yes	Yes	No
Is there a quick turnaround?	No	Yes	Yes	No	Yes	Yes
Is access to the internet needed?	No	Yes	Yes	No	No	No
Can the respondent face technical problems impeding questionnaire completion?	No	Yes	Yes	No	No	No
Is automatic processing of responses possible?	No	No	Yes	No	No	No
Can response statistics be easily generated?	No	No	Yes	No	No	No

Is there pressure on respondents?	No	No	No	Yes	Yes	Yes

Figure 6: Strengths and weaknesses of survey modes

The most discussed difference between survey modes is the lack of personal contact in self-administered questionnaires, as opposed to interviews. In a face-to-face situation (or even in a telephone interview) an experienced interviewer can tell whether the respondent is the appropriate person to answer the questions. Respondents are able to discuss issues in detail and it is possible for the researcher to explain the study in person, and to clarify points if necessary. Ornstein [1998: 68] notes that a face-to-face situation creates a 'warmer' environment which may increase respondent attention and response quality. On the other hand, personal contact may also have adverse effects on the respondents. Respondents are generally less willing to tell interviewers about things they feel will place them in a bad light than to report them on a printed (or online) form [Ornstein, 1998: 67]. While interacting with the interviewer, they may also feel the pressure to remember things in order to answer certain questions (which is not easy for everyone) and to express themselves in a way that the interviewer is able to understand.

For the purposes of the present research, a self-administered web-based questionnaire was selected as the survey mode with the greatest potential to serve the needs of this research, a mode that benefits from information age technologies and overcomes some of the limitations of traditional survey modes, while maintaining its rigour as a method of data collection. Increasingly, people are asked to complete self-administered questionnaires by entering answers directly into computers while being at work or home. Web surveys have been gaining favour with surveyors in the past decade [Couper, 2000] and this is mainly due to increased internet penetration.

A web survey has some unique characteristics that render it very attractive as a mode in the context of this research. Firstly, it appears to be the only mode that offers the possibility of conducting large-scale data collection, by accessing, as Couper [2000] asserts, undreamed-of numbers of respondents at dramatically lower costs than traditional methods. In addition to the ability to obtain quick and inexpensive responses, web surveys make feasible the delivery of multimedia survey content to respondents in a standardised way using self-administered methods. What is even more interesting is that web surveys can access almost every person with access to the internet, which means that they can reach dispersed samples all around the world, thus democratizing the survey-taking process.

A commonly discussed barrier to the effectiveness of web surveys is the fact that many people do not possess computer skills and even more people do not have access to the internet. In the case of translation professionals, computer skills are required for the

performance of most translation work nowadays [Quah, 2006: 18; Bowker, 2002: 4]; it can, therefore, be fairly assumed that the vast majority of translation professionals do possess at least basic computer skills. Translation professionals who do not use computers for their work are also less likely to be interested in TM systems, and consequently in participating in the current research. It is also a fair assumption to make that the majority of modern translation professionals are accustomed to the use of the internet and can access it when necessary – at least in developed countries – as suggested by studies of Carter-Sigglow [2004], and Fulford and Granell-Zafra [2004]. The familiarity of translation professionals with computers is therefore an advantage that is lacking in other populations that are frequently surveyed (especially when studying the general population). This, along with the widespread adoption of the internet by translation professionals, renders them an excellent candidate group for web surveys.

2.5 Challenges in data elicitation

Carrying out a survey to find out the needs of modern translation professionals with regard to TM systems does not just involve asking people what they want. The term elicitation suggests that the process is one of simple knowledge transfer where the researcher elicits and documents existing user knowledge. In reality, the process is much more complex. It rather resembles a process of problem analysis and understanding, which, according to Davies [1993], is an activity that encompasses learning about the problem to be solved (through questioning) and understanding the needs of users by trying at the same time to find out who the user really is.

Several issues may compromise the success of the process and are likely to do so unless the survey is designed in such a way that takes into account and addresses the associated issues proactively. These issues can relate either to respondents or to the quality of the instrument (i.e. the questionnaire).

One of the first issues to consider pertaining to the use of a survey as an elicitation method is that it depends on respondents' willingness, honesty, memory, and ability to respond. In contrast with interviews, where the interviewer keeps the respondents motivated and probes incomplete or inadequate responses, in self-administered surveys there is no interviewer to intervene and force the data with questions. As a result, potential problems include poor recall and communication of incomplete or incorrect knowledge, as users are unaware of their own knowledge and its boundaries, and they often forget. Further challenges occur when respondents do not feel motivated to give accurate answers, or in fact, they may be motivated to give answers that present themselves in a favourable light. Motivation sometimes also relates to reassurance provided regarding the confidentiality of

responses, in the sense that respondents may feel reluctant to answer particular questions for fear of exposing themselves or revealing sensitive information.

Drawing from experiences gained in Requirements Engineering, the greatest difficulty in eliciting information from users for the purpose of gathering requirements for a system is the fact that users often do not know what they want from the system except in the most general terms. Even when they have a clear idea of what they would like the system to do, they find it difficult to quantify and articulate this information [Kotonya and Sommerville, 1998: 56]. Similarly, not all users are able to express themselves clearly, and because of this they are often discouraged by open-ended questions. Finally, respondents may have difficulty in comprehending the questions asked in a survey or they may just as well misinterpret them, either because the questions asked are irrelevant to their particular situation or, most commonly, because of the poor quality of the questionnaire. Instrument-related issues usually include poor questionnaire design, poor wording and technical flaws.

The above issues raise a number of questions that had to be reflected upon prior to survey design. The most critical question in need of meticulous deliberation was: how can one get information on what users need, if they themselves do not know what they need? Is there a way to facilitate information flow, given that there is no interaction between the interviewer and the respondent? Other equally important questions that had to be considered were: how can the survey keep the respondents motivated in giving honest and accurate answers and encourage them to complete the survey in full? And how can the researcher be sure that the respondent understands the questions the same way that the researcher intends them to be understood?

These questions and all other related issues are addressed in the next chapter, where the survey design is elaborated. Survey design is of paramount importance in scientific research, as it determines the level of confidence in this particular data collection method. It regulates the relevance, quantity and quality of data and, consequently, affects the quality of conclusions that emerge from a survey.

Chapter 3: Elicitation of information through survey

By approaching user needs as a problem that needs to be understood before any analysis, and upon reflection on the challenges discussed in section 2.6, it becomes necessary to identify appropriate techniques for eliciting information from TM users through use of a survey. These techniques would allow the successful elicitation of rich and relevant information from users by means of purposeful questioning.

Identifying user needs is primarily a qualitative research task and, as such, it involves open-ended questions. However, simply asking users "What are your needs with reference to TM systems?" is highly likely to fail in returning the desired quality of responses and, possibly, to increase the nonresponse rate.

The commonest deficiency of self-administered questionnaires is the difficulty in accessing the tacit knowledge that people possess, due to the absence of an interviewer who can probe answers to his questions through additional questioning, especially when participants are not able to recall and communicate knowledge in the first instance. To enhance the poor predictive power of people's introspections, a survey questionnaire must include questions that cue memory recall.

A first technique considered in the context of the present survey was the elaboration of questions that drive participants to recall examples of their work situation. This can be a good starting point in the elicitation process, as questions like these define the context of

the survey and introduce participants to a topic that sounds familiar. Users are usually happy to describe their work, and they do this with ease, so that by asking them to name the tasks they perform with a TM system or the file types they deal with on a regular basis, they get mentally prepared to answer more difficult and abstract questions such as describing any functionality that would be desirable in a TM system.

Furthermore, as explained previously in section 2.5, users often do not really know, when asked, what they want from a TM system, except in the most general terms. A technique that could address this problem is the elaboration of questions that instead of asking the user directly "What would you want from an ideal TM system?" ask "Where is the problem?" or "Where are the difficulties?" with reference either to existing TM systems that users currently use, or to current work practices or routines, for those who are not using a TM system. Questions that lead users to describe the annoying features of their TM systems or to evaluate any other conventional tools they use on the job, can result in an insight into the weaknesses of existing translator-support tools and TM systems, thus revealing opportunities for future generations of TM systems.

A third technique that was considered in order to encourage the flow of ideas around TM systems, given that users often do not have a clear idea of what they would like the system to do, is to offer examples of potential features. Those features would be new – not implemented by more than 2-3 systems – and would be desirable by a number of users. The purpose of presenting such a list of tentative features would be to cue knowledge recognition and help unleash the imagination of users (which normally cannot stray very far away from the TM system that they are using).

The three above-mentioned techniques for eliciting information from users had to be tested before being applied to the survey design, to ensure that they would yield the desired quality of information from the survey participants. In the event of their proving inadequate during testing, alternative elicitation techniques would have to be found.

3.1 Pilot study

A pilot study preceded the survey design with a view to gaining insights into the main challenges of user needs identification and to help overcome some of the issues surrounding the information elicitation process. The principal aim of the study was to inform questionnaire design after achieving a triple objective:

a) Test the effectiveness of the three following elicitation techniques:

    1) asking users to describe their work routines (with or without the use of TM system) in order to familiarise themselves with the topic of study,

2) asking users to state the annoying features of their TM system (or the annoying parts of their work) so that they realise any problems,

3) suggesting potential TM system features that act as idea stimulators.

These three techniques should be able to help respondents give a quality answer to the crucial question of "What would you like to see in future TM systems?".

b) Establish the parameters (i.e. factors that influence needs diversification) that need to be measured or considered during needs analysis. The definition of parameters is important as user needs cannot be identified in a vacuum, without contextual information, i.e. without detailed knowledge about the human beings and about their life and working sphere. Needs can, of course, be influenced by innumerable factors. Nevertheless, certain types of information on the users (e.g. job role, experience, education level) are generally thought to represent the most important parameters responsible for the variation of user needs.

c) Test the instrument (i.e. questionnaire). Discuss a list of tentative questions and test their comprehensibility and clarity of meaning. Assess how easy it is for respondents to answer all the questions and if all questions are relevant to all respondents. Determine the optimum length of the questionnaire, receive suggestions for the format and structure of the questionnaire and investigate ways to minimise any misunderstanding or loss of information.

3.1.1 Design & implementation

The pilot study involved ten one-on-one interviews and one brainstorming session with eight participants. The selection criteria for all participants were fairly loose in the sense that it was only necessary for participants to be familiar with the concept of TM systems and their functionality in general terms, and for them to have an interest in them (and obviously in the research), even if they were not using any.

All participants were translation professionals and their background, credentials and attitude were checked beforehand to ensure that their opinions were credible and that they would be of help to the project. Although they were chosen at random, variety in their job roles was pursued, to allow the distinction of any differences in viewpoints, and so that each important segment of the TM user base was represented. To that effect, participants in the interviews included three translators, one researcher on translation technology, one translator trainer, two terminologists, two project managers (translation projects) and one DTP specialist, while the brainstorming session benefited from the participation of six translators, one project manager and one terminologist.

3.1.1.1 One-on-one interviews

The first three interviews were exploratory (i.e. unstructured – see also section 2.4.2.6) and were conducted in an informal setting – in the form of a discussion – with no set questions and no time constraints. An overview of the research and its aims was given to participants at the beginning of each interview, and the purpose of the interview was explained.

After collecting some basic information on the respondents themselves (such as their job title, their education level, their length of work experience and their specialisation – for translators), and asserting their knowledge of TM systems and how they work, participants were asked to describe how they would visualise the ideal TM system with reference to their needs. None of the respondents could give an answer at that point, even after making sure that they had understood the question correctly.

Subsequently, the interviewer employed the first technique of eliciting information, by using open stimuli such as "Tell me how you use your TM system in your daily work" or "Describe an ordinary process of translating". When a person had not used a TM system before, examples of TM system functions were given by the interviewer and a discussion developed on how these systems could help in the participant's job. The interviewer kept probing answers continuously, in order to keep up an interesting discussion by asking various 'how', 'where', 'what', 'who' and 'why' questions.

Participants were happy to relate details of their daily work and how they use TM systems. Especially, when asked to describe the problems they encounter when using their TM system (or any other practical problem they face with other tools), a sudden eagerness to express their complaints in detail appeared. The technique of speaking about practical problems led the interviewees to new realisations and prepared them to speak about possible solutions to their problems. Thanks to this "maieutic" method of interviewing, hidden knowledge about the interrelations between user needs and the system in question gradually emerged. Participants seemed animated and interested in speaking further, which made it possible to delve into some of the less crucial questions and to extend the interview to general issues around the research and the design of the survey. During the three interviews notes were taken of the key points of discussion.

The interviews that followed with seven additional participants were of a semi-structured format (see also section 2.4.2.6) and lasted approximately one hour each. The contribution of the first three interviewees resulted in the creation of a draft set of questions, which reflected a more clearly defined agenda for the subsequent interviews. The questions were, however, intended to guide the discussion to relevant sources of information, rather than restricting it to specific types of information. In fact, there was enough room for supplementary questions to allow for emergent themes and ideas.

The same maieutic method as before was applied to these interviews. As the users described their experiences, the interviewer kept probing, searching for better and more complete descriptions of problems and solutions. Again, notes were taken of the key points made by the participants in addition to the answers given to the questions.

During the interview stage, the questionnaire was being revised continuously through the feedback loop created by the participants and the interviewer. By the end of the last interview, the questionnaire had a refined form with clearer and more focused questions and it was suitable to use and assess at the brainstorming session that followed.

3.1.1.2 Brainstorming

The brainstorming session was conducted with the participation of eight translation professionals and one moderator (the researcher). It mains objectives were:

a) to test the third elicitation technique (do suggestions of potential TM features to the users stimulate the generation of additional ideas about desired features?),

b) to explore the factors that influence user needs

c) to arrive at a list of potential TM features that could be used in the survey as idea stimulators, and

d) to validate the questionnaire after its formation with the help of the interviews.

The session started with introductions about the research and its aims, along with a brief explanation of the motives and standing of the person who convened the group (the researcher). In focus group research, one of many possible types of bias occurs when participants, knowing the sponsor of the focus groups, feel awkward providing truthful criticisms or negative responses to their host, especially when it comes to brainstorming about a particular software product, when the session is normally organised by the developer of that product. A way to mitigate any resulting bias was for the moderator to make clear to the group that she had no affiliation with any of the TM development companies, that her research is carried out within an independent academic setting, and that her only requirement is to generate truthful and thoughtful responses for the benefit of the research. Participants were also reassured that their contribution would be anonymous. Finally, it was made clear that the aim of the session was to gather views and opinions, and that there was no right or wrong answer and no need to arrive at group consensus.

Before the discussion, a mini-questionnaire was distributed and completed by the participants with the purpose of collecting background information on each participant.

The discussion was led by open-ended, clear and concise questions (taken from the questionnaire that was formed after the interviews), neutrally presented by the moderator. Open-ended questions were mainly used because of the exploratory nature of the session

(answers to the questions were not known beforehand) and because they could help in determining which response options should be used in closed questions for the following survey. Obscure vocabulary or industry-specific jargon was avoided, so that participants could feel competent to respond to any question put to them. Good questions are believed to stimulate participants to think hard about a meaningful answer, not about the meaning of the question.

Since the focus of the session was the assembly of as many ideas as possible about desirable TM features, with emphasis on generation rather than evaluation of existing TM products, participants were told not to conform but to think the unthinkable as to what an ideal TM system could be like. Ideally, no particular TM software should be in mind when generating new ideas (which worked very well for the participants who had never used any TM system), but unavoidably users of TM systems were referring to the TM software they were using and were expressing requests that would solve the problems that they encountered with their TM tool.

It took some strong persistence on the part of the moderator to convince participants to free their imagination and express daring ideas without any consideration of the costs or the technical difficulties involved. During the session, the moderator never mentioned possible obstacles in the development process of a TM system, but instead insisted that the question in this research was not whether a particular development was possible but rather whether it would be useful for translation professionals.

Whereas at the beginning of the discussion participants' imagination needed help to be activated, after the suggestion of some ideas (mentioned by the participants of the previously conducted interviews) by the moderator, the flow of ideas among participants gradually increased. One idea led to the next, and, as discussion progressed, all participants seemed to have finally stepped into the future and could visualise a tool that would meet their needs.

The moderator also attempted an analysis of the requests made, in order to understand their needs and discover root needs. By asking "why is this functionality important to you?", "how would that feature help you?" or "why would you prefer feature A instead of feature B when they both appear to solve the same problem?", a relationship emerged between several features that were mentioned and a single latent need that no participant had explicitly mentioned. Once the moderator stated such a root need, participants were suddenly able to offer many more ideas about possible TM features around that need.

The session ran for approximately 2 hours, as the group appeared to be deeply involved in sharing various points of view and specific examples from their work. At the end of the session, participants were asked to comment on the questionnaire that would be the basis of the survey, and mark any questions that did not apply to their individual circumstances.

Finally, all participants were thanked for their contribution and were invited to promote the survey that was about to be launched if they wished for further involvement in the research. A summary of the key points, as expressed by participants, was produced after listening to the recorded focus group discussion and was distributed to all participants shortly after the session.

3.1.2 Study outcomes

The pilot study permitted a deepening of our understanding of the problems of data elicitation and pointed to the most expedient techniques for addressing them. The study also succeeded in corroborating the effectiveness of the three elicitation techniques under consideration (discussed in sections 3 and 3.1). Specifically, none of the participants was able to answer an abstract question around desirable TM functions, unless any one of the techniques was employed.

More commonly, respondents' memory needed a warm-up with questions about their working practices and ways of using TM systems in order to be able to generate ideas about desirable features. Others found it easier to discuss the problems they faced with their current TM systems and then to suggest convenient solutions. The second technique (i.e. discussing weaknesses of existing TM systems) could also be expanded to provide additional insights. By discussing not only the weaknesses, but also the strengths of existing TM systems, it gave an indication of what features are most important for users. This could be particularly useful information for developers who often have to negotiate features (i.e. give up on some features in order to add new ones, because they either do not have the capacity to develop all or simply because they do not want the system to be overly feature-rich hence complex).

The third elicitation technique proved particularly effective, as ideas suggested for potential TM features stimulated further ideas and facilitated information flow on the part of participants. In general, the pilot study affirmed that skilful questions and elaborate elicitation techniques can activate people's hidden knowledge and prepare insights for more abstract concepts such as their needs in reference to TM systems, which is the core concept within the present investigation.

3.1.2.1 Factors affecting user needs

The pilot study assisted in learning the views and ideas of a sample of TM users. Most importantly though, it helped in identifying differences between respondents, which is just as important as the views they share.

Requirements for TM features, and needs by extension, are influenced by a variety of factors, as clearly demonstrated by the fact that expressed requirements appeared to

correspond to the situation of each user. The individual parameters affecting user requirements that were identified in the pilot study are shown in the table below (Table 6). The identified parameters could be grouped into four general categories of factors: professional role & status, knowledge & experience, work environment & practices, and attitudes.

CATEGORIES	**PARAMETERS**
Professional role/status	type of translation professional employment status (e.g. freelancer, employee)
Knowledge/experience	length of work experience TM use/non-use length of TM use (if applicable) level of TM usage level of computer usage competence level of education
Work environment/practices	working languages area(s) of specialisation Operating System uses of the internet
Attitudes	perceived benefits from the use of TM systems reasons for not using a TM system

Table 6: Established parameters predictive of user needs variability

Among the most important individual parameters appeared to be a person's knowledge and experience both in the working domain and with computers. Experience and knowledge are responsible for the development of individually different strategies for problem solving, and consequently, to some extent, determine which features would be desirable in a TM system.

In addition, there is a strong correlation between the individual and his work context, as well as the nature of his job as defined by the tasks he performs in a TM system. For instance, the same individual will have different needs with regard to a TM system, if working as a freelance translator or within the translation department of an organisation. Similarly, the same translator will need different types of TM support when allocated terminology-related tasks or when asked to manage translation projects.

Requests also differed greatly between those who were using a TM system and those who were not. The latter had to be reminded (and perhaps convinced) of the benefits of using

such systems. Only when they could see some potential benefits could they generate ideas about an ideal TM system. On the contrary, if they had a negative attitude towards the system (this applied also to 1-2 participants who were using TM systems, but had a negative or sceptical attitude towards them), it was hard for them to generate ideas. This confirms the likelihood that positive attitude towards TM systems can favour the generation of rich ideas about TM systems. Therefore, the survey had to be aimed at professionals with positive attitudes or try in some way to predispose them positively to the concept of TM before asking any question on desired functionality.

Other factors that could possibly affect needs variability, such as the sex or the personal income of the translation professional, were also discussed, but no relationship could be established between these factors and any special type of need.

In conclusion, given that needs are variables affected by all the above factors, the profile of the user is relevant in order to determine the true nature of a need. The clearer picture we can get of the user, the closer we can be in our understanding of his needs.

3.1.2.2 Input for questionnaire design

The pilot study served also as a pre-test for the survey questionnaire. One of the most difficult problems for survey researchers, as observed by Ornstein [1998:2], is to design valid and reliable questions. He points out that while greater *reliability* can be achieved by systematic efforts to develop better ways to formulate questions, *validity* should rely on the respondents' judgement of what questions mean to them and not on the researcher's own opinion. During the brainstorming session, the survey questions (as formulated during the interviews) were tested for comprehension, clarity, lack of ambiguity and relevance to each professional's work situation. Participants gave suggestions for rephrasing some of the questions, and also for adding new or removing existing ones.

At the end of the session, it was also concluded that before any open-ended question on desired TM functions, survey respondents must be given a list of tentative features to stimulate their ideas about an ideal TM system. The list of tentative features should include some of the ideas discussed in the session. Fortunately, the brainstorming session produced an abundance of ideas for desired TM functionality, although not all of them could be used in the survey.

Some requests were eliminated because they were too abstract (e.g. "I'd like a more user-friendly interface") and some were considered not to concern the intended purpose of use of a TM system (e.g. automatic invoicing function). A few suggestions were also made in terms of hardware, and these were also excluded from the list. Finally, there were requests that concerned the same functionality, but were worded in two or more different ways. Such requests were rephrased into a single desired function.

From the above evidence, it is natural to expect a lot of redundancy in the answers of the survey concerning the desired functionalities. Therefore, it would be safe to aim for a large number of survey respondents to ensure a significant number and variety of original ideas that will allow the elaboration of a robust theory concerning translators' needs in relation to TM systems.

For all remaining valid ideas that emerged from the session, their quality was assessed in accordance with whether they represent either original ideas (not implemented by any known system) or poorly met requests (very few systems have implemented them). So if any of the suggested functions already existed in the majority of TM systems, based on the knowledge of the researcher or the rest of the group, they were excluded from the list.

However, even after the elimination of a number of requests, the remaining ideas were still too numerous to be included in the survey. Therefore, it was decided to include the most popular ones, that is, those requests that were voiced by the greatest number of participants. The following list of sixteen requests was produced (in random order):

"I'd like a TM system to be able to..."

- support all languages (full Unicode support)
- handle editable PDF files
- run on multiple platforms (Mac OS, Linux, etc.)
- handle graphics (extract text from images/ flash movies)
- comply with the TMX and SRX standards
- plug into a Content Management System
- support project-specific translation/ localisation directives
- have an integrated Optical Character Recognition (OCR) capability
- integrate spellcheckers for any language
- automatically sort lists into alphabetical order
- handle text in graphs and illustrations that are embedded in a document
- search the repository as I type
- show a preview of the target full text before the translation is complete
- show the full source text in original formatting while I translate
- display some co-text of the exact/ fuzzy match suggested
- construct and suggest fuzzy matches by combining different words/phrases existing in my translation memory or in my termbank

Among the less popular ideas not included in the list of tentative features were the following:

"It would be great if a TM system could..."

- integrate idiom libraries = be able to assist more in literary translation

- be able to import dictionaries from CD-ROMs
- interact with the internet (for example, by searching the internet for glossaries in one language, importing the URLs into TM bookmarks - possibility to check if those links have expired or been removed)
- have machine translation capabilities (as long as there is an option for it and depending on how well it works)
- have a TransType feature (autofill sentences during typing, according to the matches retrieved in order of fuzziness)

The resulting list was referred to all the participants of the brainstorming session for consideration and was conclusively confirmed as the list of tentative TM functions to be used in the survey.

3.1.3 Remarks on the research methods for the pilot study

The interviews and the focus group discussion proved to be two suitable research methods (see also 2.4.2.5 and 2.4.2.6) for the purposes of the pilot study. Each method's strengths were exploited to obtain the information needed prior to the design of the survey.

Interviews were more effective in exploring the issues around data elicitation, which is the reason why they were carried out first. They were also effective in providing insights into the relationship between translation professionals and the TM system and in informing the questionnaire design.

The brainstorming session, on the other hand, was successful in producing a rich pool of ideas about desired TM features. Obviously, group synergies produce more and varied user requests as each participant builds upon the ideas of the others. In fact, as it turned out, the brainstorming session proved to be a more cost-efficient method of ideas generation, as the ten people who were interviewed at first generated only ten valid requests in total, whereas the group of eight participants in the brainstorming produced over thirty in only two hours. The brainstorming session also proved to be a better means for obtaining rich feedback on the survey questions, as quite a few questions provoked a debate amongst the group over their clarity or relevance to certain translation professionals' groups.

In hindsight, fewer one-on-one interviews (especially of translators) could have been carried out without having a substantial impact on the outcome of the study. From the seventh (out of ten) interview onwards, several themes and ideas, already discussed, were being repeated and any new information referred to each person's own experiences rather than succeeding in making any new point.

3.2 Survey design

A survey's design is inherently linked to the survey's purpose. In the context of this research, the survey is used to measure the opinions of TM users (i.e. translation professionals, as defined in section 2.3) about current and future TM systems in relation to their work, through which we can identify their needs. Those needs can be then projected onto future developments in TM technology.

The survey must produce both high quality information and a high response rate, in order to allow for an accurate and inclusive description of translation professionals' needs. Therefore, its design must encourage the collection of a large quantity of high quality data.

It is generally acknowledged that the greatest part of the success of a survey lies in the design of its instrument (i.e. the questionnaire); consequently, great attention had to be paid to all aspects of questionnaire design. The section that follows describes the key considerations when attempting the best possible formation and presentation of the survey questions.

3.2.1 Questionnaire design

During the interviews (for the pilot study) a list of draft questions was produced. However, it took considerable effort to transform this list of draft questions into a set of good survey questions. The resulting questionnaire was piloted among the participants of the brainstorming session, and then modified in the light of the resulting comments (the final survey questionnaire can be found in Appendix B).

The mode of data collection (i.e. self-administered web survey) was an important determinant of how the measurement instrument was shaped. Based on this, special considerations were given to the design of a web questionnaire.

Question wording

Since the survey was to measure primarily subjective states (i.e. the opinions of TM users), it was important to focus our attention on question wording. Careful wording is essential not only because many common words are easily misunderstood or have an ambiguous meaning [Ornstein, 1998: 8], but also because certain terms and phrases have the capacity to introduce bias [ibid:200]. Therefore, it was ensured that the questionnaire contained appropriate terminology to help the respondents enunciate their opinions clearly and unambiguously and every effort was made to avoid leading questions and the introduction of bias caused by spontaneous wording. However, according to Ornstein, the respondents' interest in a topic still plays a major role in how they answer. He argues that less interested respondents are more likely to misunderstand questions and answer carelessly, while very

interested respondents pay too much attention, discerning unintended nuances in questions [ibid: 27].

Types of question

Various types of question (open-ended, closed-ended and partially closed) were used, each type selected to suit the function of each particular question. Acknowledging the fact that questions in surveys are often difficult and place heavy demands on memory or require complicated judgments, the majority of questions provided answering choices to ease the cognitive burden of responding. Especially for 'retrospective' questions about previous events (e.g. file formats used, ways of filling the translation memory database), there is strong evidence of the superiority of closed-ended questions, as suggested by the studies of Ornstein [1998:21-23]. In fact, as he points out, a checklist of answers usually uncovers some details that respondents forget when they are asked a single general question. Moreover, a checklist is also useful because not only is unassisted recall much more difficult for respondents than recognising items from a list, but the list also describes which specific items qualify for inclusion (ibid:21).

Closed-ended questions with *ordered* response categories were also used to obtain evaluative responses for well-defined concepts (e.g. "Does the content you usually deal with have a lot/average/not that much/not at all internal repetition?"). Additionally, closed-ended questions with *unordered* categories were used for many different concepts which had to be evaluated in relation to each one of the others (e.g. "If you have to translate a term or phrase and you don't know the translation, where do you usually look – in order of use?" – where the respondent had to number the resource he resorts to first, then second, then third, and so on, from a list of given resources such as dictionaries, search engines and subject experts). Many of the close-ended questions were partially closed (i.e. provided space for a different answer) to allow respondents to give their preferred answer.

All three elicitation techniques (as discussed in sections 3 and 3.1 and tested in the pilot study: 3.1.2) were implemented in the questionnaire. Before asking any broad opinion question on desirable TM functionality which would require substantial effort in first recalling information about the topic, and then in making a judgement about the information retrieved from memory, followed by additional effort in using the imagination to be able to answer such question, users were asked to evaluate their current TM systems, in the hope that the evaluation would bring to light problems that an ideal TM could resolve in future. Specifically, users were asked to rate the TM system they were using most frequently in terms of functionality, reliability, efficiency in match recall, efficiency in match precision, efficiency in speed, usability, learnability, price to usefulness ratio and customer support. A definition of each quality criterion was included in the question, to ensure that all respondents understood the criterion in the same way. The selection of

criteria was based on the quality metrics proposed in the EAGLES framework [King, 1997] and on Höge's framework for evaluating translators' aids systems [Höge, 2002], and was finalised in the discussion that took place at the brainstorming session. The session's participants concluded that these were the most important criteria that TM users consider when evaluating their TM tool, either when comparing it to other tools or when they assess its quality against their expectations or its stated purpose.

The questionnaire included two additional open-ended questions relating to the evaluation of TM systems: "Which is the feature that you love most in the TM tool that you use?" and "Which is the most annoying thing that has happened (or happens a lot) to you, while using this TM tool?". These were two of the three most important open-ended questions of the questionnaire, which would provide the qualitative data, satisfying the exploratory objectives of the present survey.

Based on the outcomes of the pilot study (see section 3.1.2), it is helpful to ask people descriptive questions about an activity before requesting evaluations of the experience, thus encouraging a more complete recall of past events. Therefore, all evaluation questions were preceded by a series of questions about the nature of the translation professional's work and the use of his TM tool(s) (see Appendix B for the full questionnaire). Recall-aiding questions related to recent events and were selected in such a way as to keep recall simple.

The third most important open-ended question of the questionnaire ("Any comments for the above possible features? Is there any other particular task that you would like to see a Translation Memory do in the future?") was placed at the end of the questionnaire as it was the question requiring the greatest mental effort on the part of respondents compared to all other questions. This question would seem fairly vague if it was placed earlier in the questionnaire, and would undoubtedly give rise to measurement and nonresponse error. Naturally, this question was preceded by questions that suggested new possible TM functions serving as ideas stimulators. Due to the importance of the question for this research, more space was provided to visually suggest that a longer answer was desired.

One advantage of the self-administered questionnaires that increases the chances for getting long and quality answers is that people can fill them out at their own speed, taking time to comprehend each question and provide a thoughtful answer.

Question grouping and ordering

There is evidence [Ornstein, 1998:23] suggesting that the order of questions can affect responses to long and complex questions, as there is a great danger that respondents will make unforeseen connections between different items in the questionnaire. The pressure on respondents to answer quickly exacerbates these effects as well as contributing to random error arising from misunderstanding and haste [ibid: 23-4]. In this questionnaire,

questions were ordered and numbered in the sequence in which they should be answered and in a way that is logical to the respondent, starting from the easy ones and passing gradually to the more difficult and abstract. The ordering was implemented in such a way that respondents' efforts to understand and resolve ambiguities in the meaning of questions could be assisted by inferences from previous questions. The ordering aimed also at reducing possible objections related to some questions (which would result in respondents quitting the survey), as it made some questions seem less objectionable in the light of previous questions already answered.

Grouping questions by topic is also important and helpful to respondents as it cuts down on the respondent effort required for focusing on each issue and giving a reasoned response. For that reason, the questionnaire was divided into seven sections, each dealing with a different topic:

1. User profile
2. Work environment & practices
3. TM use
4. TM: usage and perceptions (applicable to TM users only)
5. Evaluation of TM systems based on user experience (applicable to TM users only)
6. Limitations of TM use (applicable to TM non-users only)
7. Visualising the ideal TM tool

To define the path that respondents had to follow, question numbers and section headings were used in each section. Every section containing its group of questions appeared on a separate webpage. A 'next' button could be found at the end of each section which the respondent had to click on in order to proceed to the next section/webpage. A sense of progress was provided by numbering each section (e.g. 1/7, 2/7) in an effort to deter people from quitting when they are only a few questions from the end.

Observations of respondents answering questions that appeared one at a time on the screen suggested that some respondents lose their sense of context [Dillman, 2007: 396]. If their concentration is disrupted, even momentarily, they cannot go back and easily pick up on where they were in the mental process of responding [ibid: 396]. On the other hand, respondents' ability to see and answer groups of questions at once serves to increase the consistency among responses [Ornstein 1998: 68]. The option of having all survey questions appearing on one webpage was quickly abandoned as the questionnaire looked too long and not too attractive. To remedy a potential problem with recalling previous answers, the possibility of glancing backwards at the previous screens (by clicking a 'back' button) was provided, which would help respondents provide more accurate and consistent answers. People were guided by graphical layout features from the landing page up until the last question.

Generally, every effort was made to create a respondent-friendly questionnaire that looked attractive and encouraged people to read words in the same order as other respondents would read them.

Questionnaire introduction and closing

Responding to a self-administered questionnaire involves not only cognition but also motivation. In our case, it was necessary to find ways to motivate people to complete the questionnaire in order to achieve a high response rate.

Based on the theory of 'social exchange' [Dillman, 2007:14], actions of individuals are motivated by the return these actions are expected to bring. The theory implies that people engage in an activity because of the rewards they hope to reap, either directly or indirectly. To the same end, an effective strategy is to appeal to values shared widely by those who are surveyed, as supporting a person's values can instil a sense of reward in individuals. Our survey appealed to translation professionals on the basis of the study's "social usefulness".

The first page contained a welcome message that stated the survey's purpose and emphasized the survey's usefulness and the importance of a response from each person who was contacted. It was also made clear that the survey was conducted under the auspices of a reputable academic institution (Imperial College London), in order to engender respondent trust.

Another strategy used to motivate respondents was to make the questionnaire interesting. In general, questionnaires on topics of high salience to recipients are a powerful determinant of high response rate. Therefore, every effort was made to improve the layout and design, to order questions so that the more interesting ones were placed at the beginning, and to make questions easy to understand and answer.

The questionnaire starts with some of the most salient questions to the respondents. A salient beginning to a questionnaire is partly a matter of meeting respondents' expectations and partly a matter of identifying questions that the respondent will find interesting. Those are easy questions that require a nearly automatic answer and that apply to all respondents. Most are demographic questions (such as job title, professional status and years of experience) with obvious relevance to the topic, and highly relevant to the study's objectives.

On the last page of the questionnaire, respondents are reminded of the benefits that could be reaped by them – again, in a social exchange context – and are thanked for their participation in this research.

3.2.2 Technical features of the survey

Using the web for a survey offers many possibilities to design a creative questionnaire layout[34]. Bearing in mind how one's emotional state can influence one's ability to fill out a questionnaire, there was minimal use of colour to make the questionnaire look more visually pleasant and navigation elements such as 'next' and 'previous' buttons, as well as an 'exit the survey' link, to define the desired path through the questionnaire.

Some of the questions in the questionnaire were mandatory, so the respondent could not continue to the next section/webpage of the survey unless answering these questions (a message appeared showing which mandatory questions still had to be completed before moving on to the next section). Although this could have been annoying for some respondents, the importance of answering the specific questions was greater, justifying the use of this feature. Certainly, all the mandatory questions offered a 'don't know' or 'other' answer category (followed by a space for a different answer) to give respondents greater choice.

Other features were also used to achieve a dynamic interaction between respondent and questionnaire. A 'skip pattern' was employed – in a way that was invisible to the respondent – to direct each respondent to different groups of questions according to his answer in section three about whether he used a TM or not. Selecting the 'No' answer choice required skipping automatically sections four and five about TM use. Similarly, section six was skipped for those who did use a TM system.

Upon completion of the design of the survey, the web survey was tested for browser compatibility and then was launched online on a dedicated survey site, which respondents could access from any browser.

Each answer provided by a respondent was saved automatically in the survey data file, even if the respondent did not complete the survey in full. To prevent multiple responses from the same person, multiple completions made on the same computer were not possible by detecting the respondent's IP address.

3.3 Survey implementation

The implementation strategy for a web survey depends primarily on the selection of sources from which samples are drawn and on the alternatives for accessing respondents. The section that follows describes the sampling methodology and the considerations related to inevitable bias, as well as the process followed to recruit respondents and promote the survey.

---

[34] The survey was designed with tools provided by surveymonkey.com. It was hosted by the same company. Additional information can be found at http://www.surveymonkey.com/.

3.3.1 Sampling

Sampling is critical to survey research, as a poor sample can invalidate inferences for a given population, based on the collected data. For the purpose of this research, the target population about which the survey was used to generate a theory was translation professionals - TM users (as defined in section 2.3). Constructing a sample frame out of this population was challenging in many ways, due primarily to lack of size information or any sort of reliable directory for the population of translation professionals worldwide. The key challenge was to find creative ways to make the survey more inclusive of the desired target population, otherwise it would have to limit generalisations to more restricted populations (e.g. internet users).

The following four main criteria, in line with the aims of the research, were used to design the sample frame:

1. The survey needed to cover and elicit information from a large market share of diverse users and avoid problems of errors arising from measuring just a subset of population. Representation at the elicitation stage ensured that the different viewpoints were captured. The chosen mode of survey (web survey) helped in this direction, as we could access more people and more diverse groups within the overall population of translation professionals through a popular mass communication channel such as the internet.

2. The selection also needed to control the sample distribution geographically, making sure it was spread over the spatial dimension like the target population. In other words, the survey must reach geographically dispersed populations in as many parts of the world as possible.

3. The sample frame should yield respondents who are interested in the topic (TM systems), as these are more likely to provide more relevant, complete and rich answers than translation professionals who have no interest in TM technology.

4. Due to the fact that translators are the biggest subgroup of translation professionals who are found using a TM system, they carry more weight as a TM user group; they therefore need to be overrepresented compared to all other translation professionals' groups. In order to have data from them to prepare a separate analysis for this group, being our main group of interest, we had in effect to oversample translators.

The above criteria were used in lieu of unavailable data regarding the size of and any directory details for the translation professionals' population, which would allow us to draw a probability sample.

Our sampling method was linked to the mode of data collection and was devised to suit the purposes of the survey. Accordingly, the sample frame was restricted to translation professionals-TM users with access to or use of the web. The main reasons for this decision were: a) because it is a prerequisite for completing a web survey, b) translation professionals who use the internet are more likely to be familiar with and interested in translation technology than those who do not, and c) translation professionals who are internet users normally have an e-mail address, therefore they can be contacted easily.

The process used to construct the sample was to obtain lists of translation professionals' e-mail addresses from translation professionals' associations in every country[35], as well as from translation companies' associations (e.g. ATC). Another way to locate translation professionals was to visit translation portals (e.g. proz.com) and obtain lists of e-mails from their directories of freelancers and translation companies. In addition, the sample was reinforced by the researcher's contacts in the language services industry.

Because we needed a mechanism that could identify and involve all TM user groups, as all user groups had to be represented, a "snowballing" technique was implemented. People who were contacted were asked to pass the link of the survey to their colleagues and whoever they thought would be interested in TM systems and in contributing to this research.

Certainly, adopting nonprobability sampling methods inevitably carries various problems. Inference about generalising to the target population are generally not advisable, since it is impossible to measure statistically the representativeness of the sample against the target population (i.e. translation professionals-TM users). This happens because neither the target population nor the sample can be measured. Generalisations, in this case, can be made only on certain assumptions. The findings of this survey will represent the translation professionals' population who are also internet users, but they may also represent the population of translation professionals who are also internet users, but they may also represent all translation professionals based on the assumption that the level of penetration of the internet in the translation professionals' community (at least in the developed world) is high enough to allow such a generalisation. To counterbalance the sampling error that derives from the fact that not all translation professionals have access to the internet, we assume that those without access to the internet are likely to be either little interested in translation technology or unable to adopt it due to cost/lack of infrastructure, or simply unaware of it.

Bearing in mind that a sample of translation professionals could never be fully representative due to a lack of statistical information, based on the above criteria our sample seems to be reasonably representative of translation professionals.

---

[35] A list of translation professionals' associations was obtained from http://www.lexicool.com/translator_associations.asp and http://www.accurapid.com/journal/00org.htm

The survey is obviously biased towards internet users (for the reasons explained above) and also towards English speakers, as the questionnaire was available in English only. Furthermore, the survey is biased towards those who have a keen interest in translation technology and are willing to participate. The above sources of bias cannot have any negative impact to the survey's findings since they do not produce any bias on the variables of interest (i.e. the *needs* of translation professionals).

The desired number of responses was not specified from the beginning as no prior known attempt at a worldwide survey of translation professionals could give an indication of numbers to be expected. Obviously, an effort was made to receive as many responses as possible within a specific period, and in the event where the responses return an unmanageable quantity of data, a filtering mechanism would be devised to reduce the quantity of data to a manageable level.

3.3.2 Survey promotion

Initially, the URL of the survey was distributed via e-mail to the translation professionals and translation companies in the sample described previously (section 3.3.1). The e-mail explained what the survey was about and why it was important to respond, highlighting the study's relevance to the respondent. People were invited to complete the survey irrespective of whether they used a TM system or not. However, many of the e-mails obtained turned out to be invalid, so that the sample had to be extended to more professionals eligible to participate according to our criteria (see section 3.3.1).

The survey was additionally distributed in the same way to students of randomly selected postgraduate translation courses, in order to ensure that a sufficiently broad range of experience was captured. Apart from mass e-mailing as a method for soliciting participation in the survey, professionals were solicited using announcements on a variety of newsgroups (e.g. TRANSLATIO), announcements in a popular newsletter[36], and invitations to translation professionals' associations, asking them to pass the survey link on to their members or advertise the survey in any way they thought appropriate.

Other response-stimulating efforts included announcements in eleven TM user groups (e.g. multitrans, OmegaT, and transit_termstar groups in Yahoo!) and open invitations issued on web portals for translators (e.g. Translatum.gr). Finally, to maximise response rates, banner ads (appearing in pop up windows) were created on a translation company's website (the freelancers' section), inviting freelance translators to fill in the survey.

At a rough estimate, the survey questionnaire was probably sent to more than 5,000 people aiming primarily at collecting a rich pool of data for this research, but also at raising

[36] "The Tool Kit: A biweekly newsletter for people in the translation industry who want to get more out of their computers", authored by Jost Zetzsche.

awareness about TMs. However, the number of those who ultimately received the survey is not known, and therefore the nonresponse rate is unknowable. The survey ran for a period of 2 months (July and August 2006).

3.4 Survey results

874 people completed the survey successfully. Of those, 90% were translators and 73% were working as freelancers. The majority (89%) held a professional qualification relevant to their work and a large proportion (64%) rated their general computer usage competence as 'good', while 30% rated their computer skills as 'excellent'. 61% reported specialising in the translation of technical texts, with high levels of content repetition, and all respondents confirmed having access to the internet (which was expected due to the fact that the survey was conducted online).

So far, this has been the only large-scale survey of perceptions and opinions surrounding TM systems and it represents a unique attempt to disaggregate the relationship between TM users and systems through a multi-layered exploration of user-computer interaction. The response level was highly pleasing considering the timing of the survey (summer holiday period), and a considerable amount of data, fit for both qualitative and quantitative analysis, was collected.

Even though the survey's primary goal was to collect data for this research, its findings were sought after by various parties in order to fill gaps in the current TM technology market data. As a result, a quantitative analysis of the survey data was prepared and published soon after the survey closed.[37] With regard to this research's aims, the qualitative analysis of data is of greater interest than the quantitative, as the former will reveal the needs of the survey population with respect to TM systems.

After the survey closed, the data file had to undergo some post-survey adjustments to reduce nonresponse and measurement errors. These adjustments define the data used in the analysis step.

Nonresponse is almost unavoidable in surveys [Ornstein 1998:102], as there are inevitably people who do not complete the questionnaire fully for various reasons. Missing data can result in nonresponse error, when respondents do not answer individual questions ('item non-response') or do not complete all sections ('section non-response'). Both cases were present in our survey. With this incomplete data, the greater concern is not that we had fewer data; the level of nonresponse is quite predictable and has been taken into account in

[37] The quantitative analysis of the survey findings was presented at the ASLIB International Conference *Translating & the Computer 28* on 15 November 2006 at London, UK. The paper "Translation Memories Survey 2006: Users' perceptions around TM use" is included in the conference proceedings and has also been translated in Chinese following interest expressed by Chinese TM researchers (Chinese version at: http://hexun.com/myles/default.html [last accessed on 23/07/2008]).

survey design (by setting critical questions as compulsory in order to proceed to the next section). The more serious concern was that the data might give misleading estimates of the characteristics of the population that the sample is designed to represent.

Decisions had to be made when processing the data file about when to include a data record with less than complete information and when to exclude a respondent altogether from the analytic file.

Two strategies were considered for dealing with missing data: a) to use data from respondents for whom all the required variables were available (the drawback being that the analysis could be affected by sample selection bias, because the profile of the omitted respondents are likely to differ from those for whom all the desired variables are measured), and b) to use all the non-missing responses from the survey that are relevant to each analysis. The second strategy was ultimately chosen, as it is possible to spot a clear need even in an incomplete answer from a particular respondent.

However, not all but some of the questions in the questionnaire were considered critical carriers of data important to needs identification (see Appendix C: summary of results, questions 37-43), and people who did not answer any of these had to be excluded from the survey data file. In total, 155 incomplete responses of this sort were eliminated.

An interesting fact about the quality of responses is that, as was hoped for, a number of translation professionals who reported not using a TM system (i.e. potential users) felt comfortable in answering the question about desirable functionality in TM systems, even though they were only vaguely familiar with or lacked experience in using such systems, and contributed in that respect valuable information to this research.

3.5 Data filtering

The survey data file with the remaining 719 responses had to be further processed before the analysis to minimise another type of error found in surveys: the measurement error. Measurement error occurs when a respondent's answer is inaccurate, imprecise, or cannot be compared in any useful way to other respondents' answers [Dillman, 2007:9]. Especially, for the three open-ended questions of the survey (about the strengths and the weaknesses of existing TM systems, and about desired features), some answers presented quality issues. Not all respondents had recorded their answers uniformly, and there were a great number of ambiguous responses and a tendency on the part of some respondents to seize on peripheral issues, as well as a variation in the complexity of answers and the number of points made in a response.

All the narrative material collected from the three open-ended questions was, for that reason, processed in such a way as to filter out any non-relevant, useless or ambiguous statements.

As a result, answers to Q. 37 (about the user's opinion on the favourite feature(s) of his TM) were reduced from 477 to 442 (35 invalid statements were excluded – for example, statements like *"none"*, *"I'm really not familiar enough yet to answer this question."*, *"I don't really like it. I have to use it."*). Similarly, answers to Q. 38 (about the user's opinion on the most annoying issue(s) related to his TM) were reduced from 478 to 447 (31 invalid statements, examples of which include: *"Nothing that I can remember."*, *"very little"*, *"nothing major has happened so far"*). Finally, the 162 recorded answers to Q. 43 (about desired features) were filtered to 154 valid statements (examples of the 8 excluded statements include *"I don't use TM so far, so it really doesn't matter"*, *"none"*). None of the excluded statements were coded or analysed further.

The rest of the data was coded and organised, as necessary for a systematic analysis aiming to produce valid findings. The next chapter presents the methodology followed to analyse the collected data.

Chapter 4: Analysis of qualitative data

The purpose of analysing the likes, dislikes and wishes of translation professionals in relation to TM systems is not only to reveal underlying needs, but also to explore the broader issues surrounding the translation professional's contemporary work conditions that such needs bring into theoretical focus. An interpretive, naturalistic approach in the analysis, aiming to construct new understandings around the needs of translation professionals, is expected to turn the raw data into findings that will make a contribution to the translation technology field.

The focus of analysis is the needs of translation professionals and the context within which they are observed, whereas the goal of research at this stage is a meaningful analysis of the data collected from the open-ended questions of the survey rather than a full descriptive coverage of this data.

Coding is considered the most important process in any qualitative analysis, as it is an essential precursor to a systematic exploration and understanding of the data [Basit, 2003:144]. Coding involves assigning codes and/or categories (i.e. labels) to the 'meaning units' that the researcher identifies in transcripts, meaning units being understood as any chunk of information (e.g. words, phrases, paragraphs) in the collected data which presents some kind of interest for the current research. Basit [2003:144] contends that creating categories triggers the construction of a conceptual scheme that

fits the data. This scheme helps the researcher to ask questions, to compare across data, to change or drop categories and to arrange them into a hierarchical order. Seidel [1998] considers the analysis of qualitative data as something more than sorting, categorising and organising the collected data. He argues that analysis is primarily a process of noticing and thinking. In this process, the researcher examines the data she has collected and tries to make some type of sense out of each statement. With this in mind, the aims of the present analysis were:

- to look for patterns and relationships both between a meaning unit and the user who expressed the statement and also across meaning units themselves;
- to generate ideas that help explain why those patterns are there;
- and to make general discoveries about the needs of translation professionals.

Both a qualitative and a quantitative analysis of qualitative data were needed, as this research is interested not only in identifying needs *per se* but also in comparing these and finding out which ones are the most compelling based on the frequency with which they have been voiced by the people surveyed.

4.1 Grounded Theory

The exploratory nature of this analysis entailed the selection of a method that would allow for openness with respect to the nature and content of the results. On this basis, the Grounded Theory (GT) was selected as the most expedient method for analysing the collected data.

Grounded Theory has gained significant status in the past two decades as a mainstream data analysis approach within interpretive research contexts not only in pure social research fields but also in fields such as Information Systems [Fernández, 2005: 44, Rowlands, 2005: 81]. Originally developed by Glaser and Strauss [1967], Grounded Theory evolved over time and was later substantially reformed by Strauss and Corbin [1996], resulting in the emergence of two schools of theory (the 'Glaserian' and the 'Straussian' models of GT) with distinctive differences. Irrespective of their differences, Urquhart [2001] emphasises two key beliefs of Grounded Theory, inherent to both models: (a) the researcher has to set aside theoretical ideas before she starts coding the data (i.e. avoid preconceptions) so that she can be open to emergent themes; and (b) the concepts are developed through *constant comparison*. Constant comparison is a key process in Grounded Theory and has been the cornerstone of the analysis strategy followed in this research.

In terms of the main differences between the two models, Fernández [2005: 46] notices that these can be seen in the use of Strauss and Corbin's 'axial coding' [Glaser, 1992;

Kendall, 1999] and the form and nature of the theoretical outcome (Straussian full-description versus Glaserian abstract-conceptualisation). In the present research, the Glaserian model of Grounded Theory was adopted as most appropriate for the particular study. More specifically, the current analysis is more interested in the *conceptualisation* of data offered by Glaser than in the *full description* of Strauss and Corbin. The Glaserian approach has a strong focus on abstract conceptualisations that is not concerned with the context and characteristics of informants that are not relevant to the study's particular goal, but is tied instead to the substantive area of inquiry. This increases the likelihood of it producing a theory that is both novel and much more useful to the particular field. In addition, the Glaserian approach is far less prescriptive and offers the flexibility of a number of potential coding paradigms. Finally, a preliminary literature review on Grounded Theory has brought forward practical problems reported by researchers in using the Straussian coding paradigm [e.g. Cronholm, 2002; Allan, 2003; Sarker et al., 2001; Urquhart, 2001; Kendall, 1999], such as a loss of focus of analysis or confusion caused by the word-by-word micro-analysis of data as proposed by Strauss and Corbin [Allan, 2003].

Grounded theory's greatest advantage is that it ensures that 'interpretations are all rigorously induced from data, if the full process of constant comparison is followed' [Glaser, 1998:12]. It explicitly seeks to discover the underlying assumptions, the contexts and the experiences of those involved in the phenomenon under study, and provides a method for generating concepts systematically and with scientific rigour [Glaser, 1998:13]. Its greatest appeal for the present study is that Grounded Theory proffers the way to ground each and every conceptual hypothesis the researcher makes directly in data.

4.2 Data analysis strategy

Grounded Theory allows for various analysis strategies depending on the properties and goals of each research project, as long as they comply with the fundamental principles of *data conceptualisation* and theory generation through *constant comparisons* and *theoretical coding*. In other words, grounded theory can only be achieved when the researcher uses concepts that have emerged by conceptualising the patterns in the data (such patterns emerge when the researcher aggregates, compares and links units of meaning to each other), and subsequently by relating these conceptualised patterns using theoretical codes until a theory is constructed [Glaser, 1998:31].

In the present analysis, a key activity in the data analysis process was coding. Codes were used to assign meaning to the descriptive or inferential information compiled after the survey, and were employed as heuristic tools rather than objective representations of

data[38]. Seidel [1998:14] believes that heuristic codes (i.e. code words that act as flags or signposts pointing to interesting things in the data) help the researcher to isolate and collect the information of interest, and facilitate the discovery of things, while at the same time they help open up the data for further investigation. According to him, heuristic codes represent the things that the researcher has noticed and have the advantage of being flexible enough to change and evolve as the analysis develops.

Three types of conceptual codes, as identified by Miles and Huberman [1994], were used: 1) descriptive codes, which describe an interesting incident and involve little interpretation, 2) interpretative codes, representing an underlying web of motifs, and 3) pattern codes which represent an emergent pattern. Additionally, theoretical codes were used to denote relationships between concepts, and thematic codes were used to classify emergent concepts into themes (see section 4.2.3 Classification scheme).

Based on the Grounded Theory data analysis model, the analysis strategy followed in this research involved the following four phases.

Phase 1: Initial attempt to develop categories which illuminate the data - Open coding

At first, the documents containing the user statements were read through (see excerpts from the survey's responses in Appendices D and E), to give a general picture of the type of answers collected and to stimulate the researcher's *theoretical sensitivity*. Theoretical sensitivity, an important notion in Grounded Theory [Glaser, 1998:122-3], is what helps the researcher decide what is important to code and what is not. To enhance the theoretical sensitivity prior to coding, the researcher spent time in following advances in translation technology (attending conferences, workshops, seminars), engaging in informal discussions with translation professionals about their working practices and how they use TM tools, and in following discussions relevant to this research at translators' or TM users' fora. The theoretical sensitivity is also naturally likened to the professional experiences of the researcher, which have also helped her spot interesting information in the overwhelmingly large amount of data that was collected.

To start the analytic process, *open coding* was initially carried out. That meant that all the data was read line by line looking for anything interesting. A coding scheme was devised beforehand to guide the identification of units of interest (see 4.2.1 Coding scheme). Once an interesting incident (e.g. a stated need, a problem encountered by the translation professional, a desired solution to a problem) was noticed, it was labelled with a

[38] According to Seidel [1998:14], an objectivist approach to coding treats code words as condensed representations of the facts described in the data, therefore they can be treated as surrogates for the text, and the analysis can focus on the codes instead of the text itself. This approach imposes many rules for guaranteeing the objectivity of the code names and carries several risks (as reported by Seidel [1998:14]), such as losing confidence in the codes as some degree of subjectivity is bound to creep in.

descriptive name (code) that captured its essence in a more general way (*abstract concept generation*). Other interesting concepts that were revealing the "whys?" (why is there such need, why is there this problem, why has one chosen this solution?) were also coded.

By going through the data, more incidents were identified and more concepts were generated, which were compared to others already discovered in case they could be labelled in a way that others were labelled, preserving integrity. Concepts were compared with previously identified ones, and similar concepts referring to the same general phenomenon (e.g. a need) were grouped into a category. Categories were labelled (coded) too. Attention had to be paid in the generation and labelling of categories, so as to avoid conceptual impressions that were not carefully grounded. The categories should describe the data by means of conceptual codes, which were derived directly from the data.

Each category could have certain properties, which could be characteristics, attributes, dimensions or context. Properties of a category were also coded. Since the researcher was not initially sure which conditions of participants (e.g. professional role, experience, computer usage competence) influenced their needs, and therefore which could be relevant to each category of incidents, all available data about the profile of each user was coded as category properties (for the category containing incidents that came from the corresponding users). The relevant property items would emerge as part of the analysis in the next phase, and the others could be ignored.

After each data set, categories were revised in light of the new data analysed, as espoused by the idea of *progressive focusing* that runs throughout Grounded Theory.

Categorisation was the first step of generating a theory, as it helped conceptualise the types of needs that translation professionals expressed in relation to TM systems.

Phase 2: Saturation of categories with many appropriate incidents and further development of categories - Axial coding

The texts were read several times and constant comparison continued. The units of comparison were not the statements but the incidents observed within the statements, in an attempt to classify each incident into a category. Moving up inductively to more abstract conceptual categories, several patterns emerged through incident comparison. These patterns were also named as categories. To discover latent patterns was not an easy task as, while a few incidents gave indication of a pattern forming, they were not enough or adequately relevant to establish one. It took many incidents to be sure that a pattern truly existed.

Conceptualising and comparing constantly enhanced the researcher's sensitivity in discovery and helped to neutralise imputing while coding, since it was constantly corrected. As the statements were read again and again, categories emerged more easily.

Soon, many categories were saturated with a sufficient number of appropriate examples able to validate the category's meaning, its relation to the emerging theory and its naming.

This construction of categories led to a refocusing of the analysis. Instead of viewing the data in its original context, it was now viewed in the context of the categories. But, whilst the use of categories allows the exploration of the data and a comparison of the similarities and differences, it can destroy the bigger picture. The categories' relationships with each other were important in maintaining the overall perspective. Therefore, categories were compared with each other, leading to a reconceptualisation of the data at a higher level of abstraction. *Axial coding* was carried out, which involved making connections between categories by looking at relationships between the concepts they contained (i.e. incidents, properties, etc.).

The types of relationships between concepts could be anything from the conditions that give rise to them (causal conditions), the context into which they are embedded, intervening conditions, action/interaction strategies in which they are handled, and consequences of those strategies. This was particularly useful for identifying which of the properties containing user profile information truly influenced certain important needs. Only those that were relevant to the theory would be included in it.

The relationships were captured in theoretical codes with the intention to link the categories (and some uncategorized concepts) theoretically into a web of emerging theory. Relationships were also crucial in thinking about how categories fit together into an explanation or model, and which categories and properties earn their way into the theory depending on the links they have with other categories.

Phase 3: Prioritisation of core categories - Selective coding

In this phase, the category system was refined through *selective coding*. Because some categories started appearing to be central in this analysis, evidenced by the quantity of coded excerpts they contained, they automatically assumed the role of *sensitizing concepts*. Subsequently, moving from inductive to deductive analysis, the researcher had to go back to the documents and code once again for this particular subset of concepts and relationships, this time to find instances that might add further aspects or dimensions to these categories. Gilgun [2007:7] argues that sensitizing concepts are the ones that give direction to the researcher, and can offer a starting point from which theories can unfold.

These categories, which were seen as the most striking and interesting for the analysis, were named as core categories and were linked to several subcategories, helping in this way to keep track of grounding.

Phase 4: Overall integration of concepts in a theory through sorting

Finally, as categories and properties became saturated with appropriate coded excerpts and were linked to each other, the task of analysis was visibly approaching theoretical completeness. But in order to be able to formulate substantive and coherent theories, which are also meaningful in the context of the present research and that serve its goals, concepts were sorted into thematic categories based on a coding scheme that allowed the organisation and presentation of findings according to TM software quality characteristics (see 4.2.3 Classification scheme). While the core categories provided the larger theoretical scheme for the emergent theories, the thematic categories integrated all concepts around the core categories linking the emergent theories into a story that can find its place in the translation technology field.

Generally, Grounded Theory encourages the use of literature in this final stage of analysis in order to confirm the findings and to illustrate where the research differs from the literature [Glaser, 1998:206-7]. In this way existing theories can be 'tested' against the data. In this case, the findings from this analysis were contrasted with existing relevant theories either from practitioners or academics, in order to ensure that the interpretation offered in this research is grounded in the multiple contexts that surround the data.

4.2.1 Coding scheme

Shortly after embarking on the analysis, it was realised that for identifying user needs, the researcher needed a systematic way of spotting stated needs (if those appeared explicitly in the text) or indicators for latent needs that could be later worked out through further analysis and user profiling.

Clearly, identifying user needs in narrative material is not a straightforward task, but it is rather an endeavour of investigative or detective nature. This is mainly due to the high level of ambiguity inherent in natural language and to the fact that the same concept can be expressed in many different ways by a number of people, thus complicating and confusing the process of coding. Coding can be also derailed by a lot of seemingly interesting information contained in the user statements that bear no relevance to the research aims and that can easily distract from the focus of the investigation.

To overcome these challenges, a coding scheme was devised to define the criteria for selecting and coding units of interest at the initial stage of open coding. The scheme

predefined the focus of analysis (i.e. looking for user needs) and allowed an objective conceptual description of data. Under this scheme, the task of needs identification was based on the widely acknowledged principle in Requirements Engineering that a need, if not obvious, can be derived from a **problem** that a user is facing (with or without the use of a system) [Sidky et al., 2002] and/or a **solution** that she perceives as ideal for solving her problem(s) [Agouridas et al., 2008:25].

On this premise, the researcher was able, through deductive analysis, to spot and code both needs that were clearly expressed by the surveyed translation professionals, and needs indicators (i.e. problems, or perceived solutions to existing problems, or solutions to past problems, as those emerged from people's favourite features) (see excerpts from the survey's responses on the problems and the strengths of TM systems in Appendix D).

The task was not concerned with problems created by the TM systems themselves, as the aim of this research is not to treat the problems of an existing poor technology but is a creative attempt to contribute to the design of novel systems. Moreover, the task was not concerned with problems which, although they may result from a bad experience using a TM system, could not be relieved by even the highest quality TM system (e.g. problems that relate to poor computer specifications, laziness or lack of interest to learn or experiment with the software, or inability to comprehend very basic software instructions). Finally, problems that were too vague (e.g. "problems with opening TM", "terminology management is dreadful", "Display does not keep up") and lacked any additional information that could clarify them were not coded either.

Some examples of coding based on the above conceptual framework are shown in the following table (Table 7):

Statements (with highlighted units of interest)	Basis for coding
"I believe one of the most important features for translators would be voice recognition interface."	Explicit need
"I've just finished working on a geological report (nearly 20,000 words rendered from Brazilian Portuguese into English) for which I had to use a huge amount of my time on the Internet (Google) to make sure the appropriate wording would be used (I am an aeronautical engineer, not a geologist!). If only I could have a user-friendly tool to help me in such situations..."	Problem – Explicit need – Explicit need
"[I dislike] stored segments that are too long due to doubtful segmentation rules (e.g. a short word at the segment end which is interpreted as an abbreviation = several sentences together form a segment). Repeated subsegments in translatable text aren't found automatically although they are in the TM for sure."	Problem – Problem - Problem
"conversion of a Trados memory into Wordfast"	Perceived solution
"Search patents databases such as Espacenet"	Perceived solution

Table 7: Examples of coded units of interest and their basis for coding

A frequently discussed challenge in establishing user needs from verbatim statements is the distinction of wishes or perceived solutions from actual needs. Often, there is the danger of mistaking a perceived solution for a need [Agouridas et al., 2008:25; Griffin and Hauser, 1993:4]. For instance, a user request such as *"I'd like my TM system to comply with the TMX standard"* could be transferred to the software designer as a user need, whereas the actual need is the ability of the user (granted by the TM system) to exchange TM resources effectively and efficiently.

The reason for such a distinction is crucial to the design of innovative software solutions. Needs are abstract, generalised and technology-free descriptions of the essence of a problem, whereas potential solutions (as perceived by the users) can be, on one hand, the quick way to satisfy customers, but on the other hand, serious constraints on the creativity of software designers. In other words, needs are here to define the problem spaces. Their role is to allow the designer to understand the shape and fit of solutions which will occupy these spaces.

Certainly, the analysis and restatement of the perceived solution or the problem as a need requires investigation of the origin and rationale of the stated user request. A key element in this investigation is the contextual information associated with the user statement. This information, which was collected from answers to the questions contained in the first six sections of the survey (see section 3.2.1 Questionnaire design), ranged from data on the user's profile, her work environment and practices, her attitudes towards TM systems and her ways of working with these systems.

4.2.2 User profiling

The contextual information that accompanies each user statement plays a vital role not only in understanding the user as a person but also in elucidating different contexts of TM use, which is helpful in identifying the correct need and delimiting it within a specific user space. This information also allows the categorising of users into groups featuring the same characteristics and the establishment of links between certain needs and specific user groups.

During the analysis of qualitative data collected through the survey, various different TM user profiles were identified, some having different objectives than others, which confirms the idea that approaching TM users' needs in a holistic way is nearly impossible.

One of the most obvious characteristics that differentiated the various user groups in this research was their type of role in the translation production chain. Each user group represents a context; a context that is defined by the tasks and the responsibilities of the particular role. Examples of the most prominent user profiles, as distinguished by types of professional roles, are given below[39]:

- **Project Managers**: Within some team configurations, especially in small companies, the role of the Project Manager is merged with other roles (e.g. Administrator, Vendor Manager, Account Manager, CAT specialist) that are normally separate in larger organisations. Typically, their role involves liaising with both clients and vendors (e.g. translators) and performing standard project management tasks such as scoping the project, scheduling, budgeting, reporting the status of the project, flagging any issues to the client and assessing translation vendors. Project Managers are usually also responsible for analysing the source text against any existing resources to determine the rate of repetition and prepare quotes, and for allocating resources to the project, including distributing source texts among translators and arranging any reviewing, proofreading and DTP

[39] The task descriptions are based on the experiences gained by the author while working for several medium-sized translation companies and translation departments in a variety of roles.

stages, if required. Occasionally, their duties involve the preparation of source files (e.g. extraction of translatable text) before sending it to translators.

- **Terminologists**: The role of the terminologist is most of the time absorbed in other roles (for instance, in that of the translator or the project manager) and usually only specific translation workflows or processes call for a dedicated terminologist. Typically, they are responsible for developing terminology for particular clients/projects, extracting terminology from texts, compiling termbases, creating and maintaining online glossaries, updating and merging termbases, troubleshooting and resolving any terminology issues, and archiving and referencing termbases.

- **DTP specialists**: DTP specialists (such as web engineers and graphic designers) are responsible for the layout and formatting of material, including webpage validation or re-building. Their duties typically involve file assessment, including translatability, preparation of files to be translated, conversion of files, generation of Table of Contents and Index for the translated file and re-design of non-textual elements (e.g. tables, graphs, images).

- **Localisation engineers**: They are responsible for all technical aspects of localisation projects including project preparation, software engineering, compiling and testing.

- **TM/CAT specialists**: Also frequently referred to as Translation Engineers, they are in charge of the technical aspects of projects, where TM and other CAT systems are used. Their typical responsibilities include assessing the project files (for instance, analysis of files for internal or external repetition), preparing translation memories, batch pre-translating and marking of non-translatables, maintaining, merging, updating and archiving translation memories, troubleshooting and often training other team members on how to use TM systems. In general, they are considered the main technical resource and they are the ones who prepare and distribute localisation kits or project instructions to translation vendors.

- **Reviewers**: Reviewers are usually considered external to the production team and are normally located in the country of the target language. They are subject matter experts and are responsible for the linguistic review of the translation, including assuring its adherence to cultural and terminological standards that apply to the project. They are normally native speakers and excellent users of the target language and occasionally they may not be required to be proficient in the source language, as the focus is on the translation as an independent text.

- **Editors**: Editors are usually appointed to ensure that the content and style of a set of material is consistent and to perform standard editing tasks. An editor is normally a senior linguist who checks and edits the text for proper language, consistency and completeness, after the translation has been carried out by translators.

- **Proofreaders**: Proofreaders, who are sometimes referred to as Quality Checkers, are responsible for checking the final translation according to a set of pre-defined quality assurance steps, either required by the client or existing as part of the translation company's/department's workflow process. The role of the proofreader is sometimes merged with that of the editor or the reviewer (in small, medium-sized or low-budget projects) and involves assuring the quality of the final translation product in terms of content (correct language, numbers, terminology, style, register, formatting, etc.) and form (acceptable layout, working hyperlinks, etc.).

- **Language coordinators**: In larger organisations where several rounds of quality controls are needed, language coordinators are responsible for managing the process of linguistic quality check, software and terminology consistency check.

- **Translators**: The translator's role varies greatly depending on whether she is working independently as a freelancer or if she is working in a company as a member of a team. Again, her role can be extended to cover any of the tasks performed by professionals in the above roles or limited to very concrete tasks, such as translating straightforward texts for a specific client or in a restricted domain, and working with a specific file format, while having access to specific readily available resources (glossaries, translation memories). In many cases, translators often find themselves in situations where their responsibilities rise to include additional tasks such as converting and preparing the files to be translated, selecting the translation memories and other resources, updating the client's translation memory or any online glossaries, extracting terminology and compiling glossaries for the project/client, editing or re-designing the non-textual elements of the file, re-designing a document's layout and populating the client's translation memory. To these, several administrative and project management tasks can be added, especially if translators are freelancers working directly with clients and not through translation agencies. In some localisation companies, translators are referred to as Localisers, since their role is to translate software, its documentation and collaterals.

From the above task descriptions, a significant overlapping between translation professionals' roles is easily noticeable, that renders the boundaries of different contexts represented by each user group rather blurry.

As translators are the user group that this research is most interested in, it was important to identify additional user profiles existing within the translators' user group, so that it would be possible to determine typical needs of specific types of translators and contexts. Such user profiles were delineated by:

- the translators' level of professional experience, as a function of the years working in this role, their age and the volume of translation work;
- their computer usage competence (high-tech users vs. low-end users);
- their experience with using TM systems (people who have never used a TM system at all vs. "power" users);
- their area of specialisation (technical translators vs. translators of literature or marketing material).

Based on the above user profiling, every need was associated with a specific user group and its analysis was contingent upon the profile of the user and a particular context of TM use. User profiling proved to be very useful indeed during needs identification and helped to distinguish between different types of interaction with the system.

4.2.3 Classification scheme

Since this research looks at needs in relation to TM systems, it makes sense to categorise them, once they are identified, on a basis that is used and understood by software developers. Categorisation of needs under a classification scheme that is shared among developers is crucial for the transfer of knowledge between the present research and real TM system development projects. It facilitates the mapping of user needs to user requirements and connecting back requirements to data derived from users. Moreover, it clearly shows the connection between user needs and software quality and facilitates a seamless import of the findings to the developers' requirements modelling processes.

The classification scheme that was used was based on – but not limited to – the software quality characteristics as described in the ISO 9126 standard [1991:4.1-6]. The reason behind this was to be able to produce a theory for TM system development in the form of a quality plan, where the voice of the user can be easily translated into design parameters.

The standard decomposes the concept of software quality into six quality characteristics and 21 sub-characteristics, as shown in the following table (Table 8):

Software Quality Characteristic		Sub-characteristics
Functionality	→	Suitability, Accuracy, Interoperability, Compliance, Security
Reliability	→	Maturity, Fault tolerance, Recoverability
Usability	→	Understandability, Learnability, Operability
Efficiency	→	Time behaviour, Resource behaviour
Maintainability	→	Analysability, Changeability, Stability, Testability
Portability	→	Adaptability, Installability, Conformance, Replaceability

Table 8: ISO Software Quality Characteristics

Along these lines, the identified needs were classified and coded in thematic categories named after the above quality characteristics. However, not all ISO quality characteristics were used as no needs emerged relating to maintainability. Instead, there were a number of needs relating to the surrogate and relatively new concept of software extensibility, for which a category was created to encompass all related needs.

Coding in these six thematic categories was based on the ISO definition of each quality characteristic:

- **Functionality:** *A set of attributes that bear on the existence of a set of functions and their specified properties.*
- **Reliability:** *A set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time.*
- **Usability:** *A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.*
- **Efficiency**: *A set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, under stated conditions.*
- **Portability**: *A set of attributes that bear on the ability of software to be transferred from one environment to another.*
- **Extensibility**: *A set of attributes that bear on the capability of software to allow adding/enhancing/repairing functionality to the core system.*

Adopting industry terminology to devise a structured naming scheme for thematic categories facilitated handling and overseeing a large set of concepts within the industry context this research is interested in.

By exploring further the classification of concepts into thematic categories, needs were also associated with quality sub-characteristics. The labelling of these thematic codes was based on the nomenclature suggested by the ISO 9126 standard and the specialised reports on quality metrics for language engineering systems by the EAGLES Evaluation Working Group [King, 1997] and the European Commission (TEMAA project) [European Commission, 1997]. Quality sub-characteristics that grouped together same-type needs included interoperability (branch of functionality), customisability, learnability and supportability (branches of usability). Examples of coded units instantiating a need and its association with a quality sub-characteristic are shown below (Table 9):

Coded units of interest instantiating a user need	**Software quality sub-characteristic**
"...Difficulties in converting proprietary Trados file formats to other formats, (difficulties even in different Trados file versions)"	Interoperability
"I'd like to be able to create own file formats for import/possible modification of the available file format /customization"	Customisability
"Better Help, I have learned TM myself and it is not very easy!"	Supportability - Learnability
"...the tools themselves should be localized in more languages!"	Accessibility

Table 9: Examples of quality sub-characteristics derived from identified needs

4.3 QDAS-aided analysis

A key element of qualitative research is data management. Welsh [2002] believes that using qualitative data analysis software (QDAS) can assist researchers during the labour-intensive process of qualitative data analysis. In the context of this research, NVivo 7[40] was used to support the analysis process, following recommendations by social research scientists who have undertaken qualitative analysis with the help of NVivo [for example, Basit, 2003: 143-54] and after studying its differences with other QDAS (e.g. ATLAS) (see Comparative Review by Barry Lewis [2004]). NVivo was chosen primarily because it

[40] Further information about the NVivo software by QSR International can be found at http://www.qsrinternational.com/.

offers a range of tools for handing rich and complex data records, allowing visual coding and accurate and swift retrieval of coded information.

In fact, NVivo proved to be a powerful tool for carrying out sophisticated data coding, by offering the flexibility to code one passage in multiple ways and by supporting several ways to build theories. Its capabilities fitted well with this study's research goals and were aligned with the methodological underpinnings of qualitative analysis, as borrowed from Grounded Theory. NVivo also allowed the researcher to look at coded segments of the data in context so that it was possible to explore coded passages without separating them from the surrounding material. From within its practical interface, the researcher could easily organise the different and extensive data types and sources used in the study, and systematically store and manage the data thematically, as in the example of needs organisation in a quality characteristics taxonomy shown in the screenshot below (Figure 7). Coded needs such as 'analysis of new text against TM database' and 'separate software for managers', for instance, are shown as grouped under the subcategory (=software module) 'Project management' which belongs to the category (=software quality characteristic) 'Functionality'. The category 'Functionality' includes all those needs that pertain to the sub-categories 'Collaboration', 'Match assembly', 'Quality control', 'Resource building', 'Resource management', 'Search' and 'Translation editor'.

Tree Nodes

Name			Sour	References
Functionality			0	0
Collaboration	0	0		
compatible with earlier versio	2	2		
sharing resources - compatib	4	16		
working in collaboration	2	7		
Match assembly	0	0		
assemble (auto)	3	23		
assemble a match from exist	3	24		
machine translation	2	6		
Project management	0	0		
analysis of new text against	3	18		
control over all stages of the	1	1		
language attribution	1	1		
multiple file projects	2	9		
multiple source languages	3	6		
project management	2	6		
separate software for manag	1	1		
Quality control	0	0		
Resource building	1	1		
Resource management	0	0		
Search	0	0		
Translation editor	0	0		
Reliability			0	0
Usability			0	0
Portability			0	0
Extensibility			0	0
Efficiency			0	0

Figure 7: Needs classification in a software quality characteristics taxonomy in NVivo

The main advantages in using NVivo to aid the qualitative data analysis were:

Ease of searching for relationships: It was very useful to look at the data emphasizing the relationships within it. Using NVivo, it was easy to do cross-case analyses, to re-order the codes and generally to explore and get close to the data. The advanced features of NVivo helped to develop concepts (called 'nodes' in NVivo) and do complex thinking about the data. The sophisticated search option of NVivo, for example, allowed the researcher to explore complex ideas or hypotheses about user needs quickly and easily. Identified needs

were clearly linked to their source and the information about the user, allowing traceability.

Time saving: NVivo helped to automate and speed up many data management and analysis tasks. To some researchers, this might be the most important feature of any computer program. Most QDA programs provide tools for organising data, help shape the data in ways that permits researchers to reflect upon it, and give opportunities to see data from different angles; and all these happen in seconds.

Rigour: Overall, NVivo was very helpful in building a rigorous database for the data analysed. It demonstrated very clearly all the data coded and the way it had been coded. The relationships explored by the researcher among the data sources could easily be seen in the browser of NVivo. Welsh [2002] emphasises another important ability of NVivo to add rigour to the qualitative studies: a search facility that enables researchers to interrogate their data. This ensures that the theoretical ideas that have emerged in the first round of coding can be systematically evidenced in the data, thus addressing testing the validity of the research results and making it easier to see if all data relevant to, for example, "lexical resources sharing" are coded electronically together rather than manually highlighted on paper. This, in turn, makes it possible to search this "lexical resources sharing" code to find out, for example, which users are concerned about this issue, what problems/conditions gave rise to their need and which needs are similar or conflicting to this one.

However, NVivo presented also some drawbacks which started appearing as more data was imported into the system and coded (the total amount of qualitative data imported into the system was approx. 22,000 words). The most serious of these has been the distortion and sometimes loss of highlighted quotes included in a code or category. Because of this malfunction, each code and category had to be compared manually against the records of qualitative data to make sure that the quotes included were full and that no other important quotes were left out. This procedure caused a significant delay to the process of data analysis. Other drawbacks included slow response time when querying a segment of text or a respondent ID, and frequent 'no response' errors.

4.4 Challenges in data analysis

One of the weaknesses of surveys, common in all data elicitation methods that have natural language as output, is that they require coding, which is error- and bias-prone. Difficulties associated with coding include: a) Can the codes be interpreted the same way by other people? and b) Can the researcher avoid being influenced and resist applying concepts from elsewhere?

It is commonly known that natural language is ambiguous and open to misinterpretation. It is, hence, possible that people from different backgrounds will offer different perspectives when called to interpret the same data. Griffin and Hauser [1993], acknowledging the fact that interpretation of data can never be free from some degree of subjectivity, suggest that the ideal scenario for analysing data in order to produce user requirements consists of a multi-disciplinary interpretation carried out collaboratively by analysts from different backgrounds. This variety of perspectives is thought to lead to a larger set of user needs and a richer understanding of the user than is feasible with a single expert [Griffin and Hauser, 1993:12].

Since a multi-disciplinary interpretation of data was not possible in the context of this project, the means to reduce subjectivity and enhance the interpretation of needs rested on the analysis strategy employed by the researcher. Specifically, subjective bias was kept to a minimum, by adopting the following measures:

- The researcher used a methodology (Grounded Theory) capable of providing a systematic way to control bias: by comparing units of interest to each other, any patterns that emerged and could contain bias would soon be corrected by further comparisons between units of interest, categories and their properties. Indeed, some distortions in identifying patters were revealed and corrected as the comparative process progressed. Similarly, the meaning attached to a segment of data was always tied to a specific context, for example, the profile of the user, which minimised the need for imputing.

- Scheme-based coding was another way to mitigate the risk of subjective analysis, by establishing a systematic way of objective conceptualisation of data based on a predefined conceptual framework. This way, the researcher's decisions about relative importance in the data were based on fairly objective criteria, rather than being guided exclusively from her theoretical sensitivity. Moreover, the grouping of concepts under a classification scheme that is shared among requirements engineers added further trustworthiness to the conceptual architecture of the emergent theory.

- Subjective interpretation of data can also be minimised when an alternative explanation for a puzzling finding is offered from the literature. In this research, findings were compared to existing information available in the practitioner's domain, in order to ground the interpretation in multiple external viewpoints of the phenomena in question.

- Sometimes field experts have preconceived notions, which cause them to miss surprising or unexpected statements of needs. In this research, it was important to maintain an open disposition and a willingness to be surprised. The researcher

periodically stepped back and asked 'what is going on here?', especially when contradictory concepts seemed to exist for the same phenomenon.

Inevitably, the findings have been shaped by the constructs and experiences of the researcher conducting the research and carrying out the data analysis. It was impossible to avoid influencing the coding with the researcher's store of general knowledge of research and theory and personal and professional experience. On the other hand, this helped to produce a unique picture of needs in relation to TM systems that benefits from the researcher's personal experience and skills. In any case, however, it was one of the researcher's foremost priorities to try and maintain an interpretive style balanced between a personal hermeneutic perspective and an objective conceptualisation.

Chapter 5: What are the actual needs of TM users? – Part I: Functional aspects

User needs, as identified within the present research, refer to both functional and non-functional aspects of a TM system. The functional aspects concern the functionality of the system, i.e. the operations that it is required to perform. These aspects are presented and discussed in this chapter. Non-functional aspects, on the other hand, include usability and utility issues, for example user interface design, system response times, search results retrieval time, match recall and precision, portability, and software resilience. The discussion of these aspects follows in Chapter 6.

It is important to stress again that needs should be expressed at a high level of abstraction (i.e. stated in a way to avoid implying a particular technical solution – although suggested solutions are acknowledged), leaving room for a broad range of possibilities and alternatives to be considered while designing a system, with the designer in a position to generate new and innovative features with these descriptions. It is also important to note that the determination of needs, in order to find meaningful TM application fields, must be examined with concrete reference to use contexts, as users' needs cannot be properly understood without an awareness of what users are continuously exposed to.

5.1 Functionality

According to the survey, one of the most frequently expressed observations by TM users appears to be that TM systems are strongly domain specific, namely they cater for the needs of translation professionals working on technical translation projects. The importance of technical translation in this day and age is unquestionable, as it is estimated to account for 90% of the world's annual translation output [Kingscott 2002:247]. There is therefore an obvious overwhelming need to support the translation of technical texts that come in great volumes and involve a degree of repetitive, tedious and often sterile processing.

However, this statistic does not mean that 90% of all translators are specialised in a particular technical field, spending all their professional life translating texts from this field. It is even possible that there are hardly be any translators of this sort, as people (and, indeed, translation companies) tend to seek variety in their work, and often grasp any opportunities for profitable jobs in areas outside their main domain of expertise. This means that Translation Memory tools that rely on the principle of "you will never translate any previously translated text twice" (the slogan of a well-known TM software brand) do not seem to produce benefits for a substantially large segment of translation professionals, not only those who translate non-technical texts but also all those who *"translate in a variety of technical fields, sometimes all different throughout the year, hence repetition is not a major factor"*.

In fact, the survey reveals that a significant number of respondents (33 out of 145) who are not using a TM tool believe that TM systems are not suitable for their work because of the types of content they deal with (e.g. film/series subtitles, literature, marketing material, food, tourism, culture, historical and philosophical texts, musicals, legal, market research) or because they are simply generalists. A number of translators have also expressed frustration and disappointment when trying to apply TM technology in certain contexts.

As for the actual TM users, they too believe that TM systems, as they stand, are not suitable or useful for all content. Quite a few of the questioned TM users reported refraining from using a TM system in cases where their work does not contain enough repetitions (internal or external) to warrant the use of TM tools, as they sometimes take on *"translation jobs out of [their] regular fields of expertise"* or they *"do not always translate for the same company or about the same subjects"*.

Some TM users present this case of content variation very clearly:

> *"...I only find TMs useful for a reduced number of texts, i.e. user manuals, software, etc. When it comes to scientific texts (academic papers, legal documents other than forms and the like, advertising texts, even technical texts with little repetition and metaphoric, somehow literary or colloquial language), I do not think TM tools are useful at all, on the contrary, I think that segmentation equals 'text grinding' and lack of cohesion in those cases."*

"Since I translate very technical documents from many different fields, the specialized vocabulary doesn't overlap much from one project to another except for the same client."

"[There is] not enough repetition in some of my texts and even between texts. I receive a very wide variety of texts from one of my clients and may work in the same subject area only once. For my marketing clients, the repetitions are primarily at the subsegment level, which limits some tools."

"In general, the scientific translations, which are a large part of my work, are poor in repetition (internal or external). Sometimes, in those texts, the author uses too large sentences, in which segmentation can make the work difficult."

Content repetition is widely assumed to be high in technical texts, which leads TM software companies to promise high leverage of existing translated content, and consequently of translation cost savings, to prospective buyers. But, as shown above, this is rarely the case for many translators. Research conducted by MultiCorpora Inc. points to the same conclusion by reporting that "complete sentence repetition is almost non-existent in all types of content except technical manuals for products that belong to a large family of similar products". In addition, they maintain that "in all cases of technical content, the majority of repetition comes from expressions of 5 words or less" [Tardif and Laugesen, 2006].

The reported limited applicability of current TM systems reveals an immensely large area of translation activity not being able to benefit from the existing TM technology. As appears from the survey's responses there is a strong need for tools and techniques that can increase translators' productivity without the benefits depending on the repetitiveness of the text. The need to expand the scope of use of TM systems lies behind this realisation, as well as behind the attempts of many translators to apply the existing technology to their work with disappointing results. Scope expansion can be achieved with such functionality that will allow the processing of all the existing content types that come in electronic format (e.g. content from DVDs, software, games, small screen devices) and the provision of considerable assistance for translators who translate any variety of non-creative texts (containing standardised expressions and terminology), as well as generalists and even those who engage in creative translations and need some help from time to time.

In addition to this general need pertaining to the functionality of TM systems, a number of surveyed TM users expressed the need for TM systems with broad language support, after failing attempts to find or efficiently use a TM system for their language combination. Full language support does not only mean true Unicode software, but also its ability to display correctly all languages supported by the operating system of the computer. Text with extended (non-ASCII) characters should be rendered appropriately, including ligatures, composite characters, writing direction and so on. This is an unconditional need as the fruits of technology should be enjoyed by all.

Apart from the above general needs calling for functionality enrichment in a way that could expand the scope of use of TM systems, the survey has revealed a great number of specific needs concerning functionality, which correspond to TM software modules as inferred and grouped from the specific needs expressed by surveyed users. These modules concern the translation editor, resource building and management, search, translation assembly, project management, quality control and collaboration.

Modules like these are present today in many of the available TM software packages and help to organise the various functionalities that each TM system offers, as well as facilitate the access to a particular set of tools by a particular type of translation professional who only needs to perform certain sub-tasks of the full translation process (for example, a terminologist can open the module that contains the resource building and management tools without being overwhelmed by unnecessary tools used mostly by a translator or a project manager, which form instead part of the translation editor or project manager modules).

These modules need to be distinct and operate independently (although, it is essential that they are linked dynamically and that they are part of the same application), so as to avoid imposing a specific procedure for processing a translation project (for example, some translators want to be able to start translating a text right away without setting up a translation memory or termbank beforehand; others, prefer extracting terminology as a first step before embarking on translation). Obviously, each project, depending on its volume, its type and the requirements of the client, has different needs in terms of the tasks required, and the same is true for the needs of each translation professional working on the project and performing different tasks, often in his own individual way.

5.1.1 Translation editor

The needs that concern the functionality of the translation editor are different from user interface requirements (to be discussed in section 6.1.1). The former relate, above all, to the translation process that the user enjoys or finds more convenient to follow while he carries out the translation task.

The survey showed that translators have different preferences in terms of the type of translation editor, depending on their personal circumstances, and also sometimes on their idiosyncratic work habits. Based on their responses, translators choose one of four ways when it comes to translating with the help of a TM system; consequently, needs associated to each one of them must be examined separately.

1. Some prefer to work in a **standard text processing environment (e.g. MS Word)** that they are familiar with, and for which TM plug-ins have been developed to bring translation-assisting functionality into these programs. All users who

responded favourably about these word processors were receiving the greatest proportion of their work in an MS Office file format. One of the commonest reported advantages was *"...the text is translated and the doc's layout remains the same as it was it in original. No need to redesign/re-layout the output"*. Some favourably inclined users liked the physical presentation of segments and others appeared to be particularly fond of working directly in the environment they were accustomed to, having the ability to use MS Office's rich text editing functionality (*"so I can use Autotext and even macros, as well as spell check 'on the fly'"*). Furthermore, a repeated comment coming from translators and translation company owners alike was that an editor of this sort is easy to understand, hence it *"doesn't freak out new translators as they still can work in Word"* and *"doesn't bother [translators] with the interface that has to be studied"*. However, the majority of the favourable users of this editor reported major problems with it, the most repeated of which was the frequency of MS Word crashing, *"especially with documents that contain odd formatting changes"*. Another problem worrying some of the users was the frequency of updates to the RTF file format by Microsoft, and the inability of TM developers to provide plug-ins for this format to keep up with these changes. Translators using this type of editor expressed a need for correct language detection, the ability to merge and split translation units and that the TM application should *"not disable any functions of the software for the source file (e.g. when translating .doc files, all MS Word features should be available)"*. Moreover, being able to use Copy and Paste in the MS-Word environment while using the TM tool appears to be also important for these users.

2. Others favour working in a **dedicated text processing environment** into which the file to translate can be imported. The advantage of a dedicated interface is that as long as the translator receives work in any type of format supported by the TM system, he will always work in the same interface in the same way. Editors of this sort protect the original formatting by default, and this information is also applied to the translation. However, unlike word processors, the editors make it much more difficult to modify the formatting information, and this was the most frequent complaint expressed by the users of these editors. The ability to copy source segments and paste them to target segment cells, as well as that of being able to merge and split segments, all at a touch of a button, were the next most compelling needs of these users.

3. Some like better working in a **translator-friendly word processor** to which one can copy and paste any text from any file into the application, translate it with potential matches, and copy it back into the originating application. The translator-friendly word processor has all the standard word-processing facilities enhanced to facilitate the translator's text editing moves. People who use such editors enjoy the

benefits of TM functionality irrespective of the file format of their source file and are in greater control of the page layout and the formatting of text which they can edit or leave as it is from within the native application. A further advantage reported by the users of such editors is that they can choose which text needs to be translated and what needs to stay as it is.

4. Finally, others prefer translating in the **native application of the source file** (e.g. Adobe FrameMaker, InDesign, Photoshop, Dreamweaver) **or within a CMS**, and have their TM software open at the same time giving them access to stored previous translations or terminology. There were quite a few cases like these present in the survey, involving work scenarios such as translating creative material (*"when translating advertising/marketing material, it is sometimes more convenient and inspiring to work directly on the actual source text. TMs sometimes make me translate mechanically, word for word; this is more applicable to technical texts rather than creative translation."*), texts with complex layout (*"Sometimes it is easier for me to translate the text 'manually' and then perform alignment of the source and target texts in order to avoid formatting problems if the text seems to be too complicated from the formatting point of view."*), short jobs that *"don't always justify the work of setting up a project in the software's data organization system"* or translation of minor updates (*"If there are small updates to projects, it is easier and quicker to implement them directly in the source files instead of going through the whole rigmarole of converting files backwards and forwards which can be quite cumbersome..."*). Some of these translators requested a minimal presence of a translation editor (as a plug-in) in the form of a toolbar lending translation editing functionality to the text processing application of their choice. The TM functionality would be accessible from the toolbar which would be always on top of the window containing the project a translator is working on in any file format, so that the TM functions are always within easy reach without impeding his work. The translator would perhaps highlight a term, phrase or sentence and by pressing a shortcut the system would search the databases/index for this highlighted segment, displaying the results in a small search window. This design appears to be ideal for those who want minimal help, especially with terminology, who prefer greater freedom in translation and who do not encounter great external repetition.

Due to the fact that different translators follow different translation processes, an ideal TM software intended to assist several translators working on the same project or for the same company would be able to offer a flexible text editor, either by giving the translators the option to use any of the above alternatives, or by combining them into a new and versatile design.

5.1.1.1 Translation process

In terms of the translation process that each editor offers, the first two models described above impose their own way of working through the text, with little flexibility given to the translator for customising the process. Most of these two types of editors segment the source text, forcing the translator to work on a segment-by-segment basis. The feelings of the surveyed TM users about this rigid process were mixed. Those who usually translated creative or non-creative, yet not technical, texts maintained that the segmentation of the source text:

> *"...hinders the flow and creativity of the translation process due to the 'one sentence/one segment' structure."*

The same group believed that this process can also have a negative impact on the translation product:

> *"Forcing me to work on a sentence-by-sentence basis interrupts the flow of the language, and possibly degrades the quality and beauty of the language. Granted, this isn't a huge issue with highly technical texts, but once outside the realm of purely technical, style DOES matter."*

However, a few inexperienced technical translators, having Romance languages as their working languages (e.g. French, Italian and Spanish), expressed a positive view about this process, when asked what they liked most in their TM software:

> *"[I like] that it forces me to work steadily and consistently through a text. Before, I used to 'jump' difficulties and leave them to solve later on. With [name of TM software] I have to work each one as it shows up - this is very good."*
>
> *"[I like] the physical presentation of segments (in [name of TM software], in two differently coloured boxes one above the other, with differently coloured backgrounds)"*
>
> *"[I love] the fact that you can translate a sentence at a time without seeing the rest of the text. This makes it easier and clearer to translate."*

For this group of translators, the major concern is the ability to change the order of translation units, which they need in order to preserve their freedom in manipulating the structure of the translated text.

Others, apparently not satisfied with the existing process, expressed the need to be given *"the possibility to go paragraph by paragraph seeing the source and target segments"* while they translate, and even the option *"to choose how many words the segments I'm translating should contain"*.

In literature, segmenting the source text into sentences has given rise to numerous objections from translation theorists and practitioners alike, who claim that such a process induces to a mechanical way of translating which prevents the translator from applying any creative flair to the translation and even alienates the translator from his work [Biau Gil, 2005; Bowker & Barlow, 2004:79]. Pym [2003], on this matter, insists that "seeing texts in

segments reduces translation to the most primitive sense of fidelity: fidelity to words at sentence level", while he argues that such a substandard process "keeps translators focused on language-replacement exercises". Désilets et al. [2006:22] add a further argument against this process by explaining that:

> "...it tends to encourage (sometimes even enforce) literal, sentence by sentence translation. But good translation which is idiomatic and culturally appropriate often requires that the translator deviate from the structure of the source text. In particular, whole sentences may be left altogether untranslated if they are not culturally appropriate in the target language".

Overall, the surveyed TM users seem to prefer being asked how they want to work through the text: sentence by sentence, paragraph by paragraph, or seeing the whole text. This way, the system will adapt to the translation process that is most suitable for each translator, rather than imposing its own.

A side-effect of following a segment-by-segment translation process is the difficulty of seeing or visualising the source segments in their original place in the source text, as the text no longer exists as an integral semantic unity but has been dismantled into small individual textual fragments:

> *"With [name of TM software that allows working in MS Word] you cannot see the full source text. Once you translate a TU the source text disappears. Same with [name of TM software that provides a dedicated text processing environment]: You cannot see the entire source document. You have to open it in Word to see what it looks like in full."*

This causes several problems and limits the application of TM systems, as some translators have reported:

> *"It takes text out of context. Sometimes it is difficult to recognise and maintain format."*

> *"I do not use TM software for literary, journalistic or creative texts, that is, for texts for which I need to have the whole text in mind."*

Along similar lines, Benis criticises this particular translation process in one of his regular TM system reviews [2007a:30] by explaining that:

> "...you see little context above and below the sentence you are working on. More importantly, when the time comes to check your work you will find the sentences displayed contiguously, making it easier to lose your place or inadvertently skip a sentence, and more difficult to scan through your work, checking the translations against the source text."

On the issue of source text visibility, using frequency of mention as a surrogate for importance, the survey showed that the visibility of the source document at all times (*"the capacity to work side-by-side (Source-Target)"*) during translation is imperative to any translation process. In fact, it appears that it is not only enough to be able to view the source text as a list of sentences, but it is also important to see the formatting of the text, as well as any non-textual elements, in other words, to be able to view the source text in its original form. This need appears to be associated with the translation of all types of texts, as

well as with the localisation of software. With regards to the latter, Savourel [2001:17] points out that the lack of source text context during the localisation of software poses "a restriction that can affect the speed and the quality of their work"; for example, "a heading can be translated differently or have different capitalisation rules if it appears in a manual or in the index of a help file".

Moreover, research on the importance of visual elements that are present in the source text has shown that it helps translators in problem solving and decision making during the translation process [Kussmaul, 2005]. In fact, Kussmaul [2005:378-82] believes that the visual elements of the source text aid the comprehension of the text and, furthermore, they help stimulate a target frame, that is a translation. Biau Gil [2007] has also observed, through his experiments, a difference in the quality of translation when translators had access to visual information compared to others who did not. He argues that the lack of visual information in the translation environment creates problems, such as wrong punctuation, inadequate translation decisions due to a lack of awareness of a problem and delay in decision making.

Another issue brought forward by a small number of surveyed TM users and relating to translation editors and the translation process that the system dictates, was the lack of support for the reviewing process which is required from translators/reviewers as a service as often as translation itself:

> *"I also regret that TM does not fit well into workflow when freelancers are editing/being edited. The final changes do not usually make it into the TM, since cleaned documents are sent for editing."*
>
> *"It is very important to have an environment for the reviewer, for example, identify only the segments to be reviewed and automatically open only these segments (e.g. sometimes a reviewer doesn't have to review 100% matches)."*

For this group of TM users, the translation editor needs to incorporate functionality that is appropriate for reviewers, such as the possibility to:

- select the text that needs to be reviewed
- track the changes that have been made
- accept or reject changes
- add comments to translated segments
- do a global search and replace for any text
- present to the reviewer, once he corrects or edits a word/phrase, all the occurrences of this word/phrase throughout the target text, in order to correct it in all other instances, as well as all the occurrences in any of his chosen resources, in order to correct it there too
- compare source and target texts and highlight differences between the translation and the user's translation resources
- use all the quality control tools (see section 5.1.7) that are incorporated in the TM software.

On the whole, translators showed great interest in working in a text editor that allows them the flexibility to translate creatively and in their own preferred ways:

> *"I don't want the CAT tool to force me to work in a different way so I like to continue doing some work without it to make sure I keep my usual translation methods."*

The more experienced the translators were, the more freedom they appeared to demand in their text-editing moves. As for the translation process, all translators seemed to enjoy a process where the system automatically displays the translation of any source text that seems to have been translated before, as the translator works on his text, having at the same time the possibility to edit any text in his repository of language resources (e.g. past translations, glossaries, corpora) interactively as he works.

5.1.1.2 Target text preview

Another important need frequently mentioned in the survey responses is the ability to *"preview the target file, no matter how far the translation has gone"*. In particular, translators appeared to appreciate the possibility of previewing their translation in its original file format at any time by clicking on an icon, in order to ensure that they have not missed any non-editable text that also requires translation.

5.1.1.3 Visibility of formatting tags and underlying code

According to the survey, one third of all electronic editable text for translation comes in formats other than the MS Office ones (e.g. HTML/XML files, InDesign or Quark files). Translators dealing with these types of texts appeared to be concerned with the markup instructions hidden in the text, which, if deleted or corrupted by translators by mistake, cause problems such as the damage of the formatting of the text or the layout of the document. In some other cases, the offending codes may even prevent the exporting of a project from the TM application.

Bowker [2002: 137] argues that the ideal TM tool should "shield translators from the messy business of wading through computer code and allow them to focus on identifying and translating the relevant text strings. The advantage of this approach is that it helps to reduce the translator's learning curve, which makes the translator more receptive to using the tools." As is indicated by the survey, her view is shared by a particular segment of translators, i.e. those with a low level of computer skills and with less than five years of professional experience in translating. These translators need protection from something they believe can disrupt their work, and therefore it should be left to specialists to deal with. All other translators with good or excellent computer skills, irrespective of experience, TM usage level or age, seem to prefer the freedom to interfere with the code and change it, in order to adjust the formatting of the text or the layout of the document. These translators favour the idea of a WYSIWYG editor that can display for both the source and target files

the formatting and layout and also enable the user to access the source of the files where markup is visible and editable, with the additional possibility of copying tags from source text to target text. This would satisfy a compelling need expressed by the surveyed high-tech translators to *"see the 'bones' of a text"* and to be able to manipulate the form of their translation.

5.1.1.4 Pre-translation

On frequent occasions, translators receive partially translated texts and are called on to translate only the untranslated parts. This is often the case when there are regular updates to content that needs to be fresh at all times, such as product documentation, some websites and financial reports. The possibility of carrying out a pre-translation of a document in a TM system, by leveraging any previously translated content, appears to be very appealing to a large number of the survey respondents.

The advantages of this function are found to be appreciated by both translators and translation clients alike. The latter choose to use pre-translation when new text has been added to an old translation, or when they are reluctant to assign anew the translation of text that has been translated before, especially when they can find the old translated segments in their TM repository. Apart from the obvious cost savings, many companies/clients also prefer to supply external translators with pre-translated texts, instead of giving them their translation memory, to protect the ownership of their database. Finally, by providing a pre-translated text to translators, which contains the terms, phraseology and brand names used throughout their organisation, they ensure that their branding is preserved. In all cases, the translator's job is to translate the remaining parts of the text for which no matches could be found in the client's repository.

Translators also appeared to be in favour of a pre-translate function, under certain conditions. If their client has carried it out in advance and has provided them with a partially translated text from their resources, they appreciate having the terminology, phraseology and style that the client has already approved right in front of them, saving them time to research such matters. Things change, however, when the pre-translation has been carried out in a TM or MT system, which has produced a draft translation by machine translation techniques. There is no doubt, as Wallis has also rightfully noticed in her experiments [2006:12], that most translators do not like working as post-editors for these kinds of (partial) translations. And that is because, first of all, it is generally more time consuming to edit a poor translation than to produce a new one and, secondly, with pre-translated segments the translator is inclined to use the sentence structure of the source text (particularly problematic for languages with hardly any similarity between their sentences' structure), with little freedom or motivation to apply his own style, while at the

same time risking producing a 'jigsaw puzzle' with pieces that seem to fit correctly with each other but fail to form an overall picture.

It appears that pre-translation is a helpful function for translators when clear-cut specialised terms are replaced by target language equivalents at the beginning of a translation. Most problems start to appear when the system automatically replaces every part of a source segment with the target match (fuzzy or exact) that it finds in the database. Therefore, it makes sense for translators to be given options as to which matches he wants replaced. To the same end, another solution expressed by the surveyed translators would be the possibility for them to carry out a pre-translation interactively, that is, to enable them to accept or reject the proposals of the system at their point of entry, rather than after they have replaced the source text. This way, the translator can control any unnecessary replacements before starting the translation, rather than spending time editing them at a later stage.

An interesting need inferred from the comments of translators regarding pre-translation is the possibility of viewing the pre-translated text not in the main working window but in a separate window, using it as a reference, instead of a starting point for their translation. Some translators have actually reported that a pre-translated text is often so restrictive that they cannot work directly from it but prefer to create a brand new target text; yet, it is often so helpful that it can be used as one of their primary sources of consultation.

5.1.1.5 Preservation of text formatting and page layout

One of the major concerns among TM users who use either a MS Word-based editor or a dedicated text processing environment seems to be how well the system deals with complicated formatting and document layout. A great number of complaints were expressed regarding the editors' propensity to damage documents with complex layout, *"causing formatting problems that call for repeated document re-creation"*. Most problems are encountered when the source text contains tables, numbered sections, bookmarks, bullet points or frames. In addition to altering the formatting, some systems also fail to recognise translatable text within text boxes, footnotes, tables of contents and links, which means that this text is often not presented to the translator for translation, hence translators often forget to translate it and only realise it when they preview the final document – at which point they go back and sometimes translate it outside the TM.

For the users of these types of editors, the system's ability to preserve the original formatting of the text and the document's layout in perfect condition, as well as identifying text in complex formatting appears to be of paramount importance, as the lack of these features is causing them great frustration and delay in their work. Additionally, some also favour the possibility of cloning the formatting and layout of the source text to the target text, arguing that a perfect cloning can save them time in recreating a complex formatting.

Nonetheless, it is agreed that the editor must allow the user to change the formatting and layout as he sees fit and must offer a range of formatting options.

For the users of the third type of editor (translator-friendly word processor), no requests or comments were reported regarding formatting issues, apparently because the translators who use these editors can adjust the formatting and layout in the native application of the document they work on, without the TM system having to deal with any of this – the translator just copying and pasting plain text in the TM editor.

5.1.1.6 Productivity enhancement tools

Apart from the previous needs regarding the functionality of translation editors, the surveyed translation professionals showed interest in various tools integrated in an editor that they believe would enhance their productivity during translation. The most popular ones were those that would enable users to:

- sort lists alphabetically;
- apply different locale-specific styles easily, according to the rules of which:
  - ordered lists or tables can be sorted,
  - automated text, such as quotation marks or captions, can be generated,
  - measurement units, dates and time can be converted automatically;
- monitor the translation progress by real-time word counters and progress bars (for example, a user has proposed a small field on the status bar that shows how many words the current document has, with the number being updated as the translator types. Another user has proposed a small counter showing an ongoing count of how many words/lines there are to go in the file or project, complete with match analysis); the word counter is expected to be accurate and it is important that it does not miss out any hidden text (e.g. alt descriptions of images) or text in headers, footers, text boxes, etc.;
- propagate matches throughout the document on the fly, with the user being allowed to specify the type of matches (exact, full, fuzzy, terms only, or from a specific resource) that he wants to propagate;
- perform any kind of global replacements (e.g. replace commas with full stops in figures), applying grammar rules where necessary (e.g. changing the form –plural, singular, feminine, masculine– of nouns and adjectives according to the co-text);
- copy and paste text from other files that a user may open, or webpages;
- copy text from source text to target text;

- add a comment to a text segment (e.g. a query by the translator on a particular term/phrase);
- incorporate terms from a terminology database more efficiently, for example having a letter assigned to each term in the database which can be entered with one keystroke as the user types;
- offer a space for notes of queries and uncertainties on the translation in progress – the notes must stay there on screen, in context, reminding the translator of the problem until he has solved it and deleted them.

Great interest also emerged in an interactive mode of translating where the user by starting to type his translation automatically triggers off the matching engine, so that matches are invoked by a keystroke. As the user types in his translation, the system then dynamically predicts alternative translations that best complete the part of the sentence being translated, based on the retrieved matches. Separate areas of the editor can present all the different matches. This autocomplete feature is believed to speed up writing time by limiting the number of keystrokes and avoiding mistype errors, as long as it allows for adjustments, i.e. allowing the user to specify the types of matches that he wants to allow autocompletion for. Some users expressed their concern about the system's interference with the typing which can be very irritating when the matches suggested are not the right ones. In such cases, the user must be able to de-activate the feature or set it to allow the autocompletion of exact matches only.

5.1.2 Resource building

Translators try constantly to build and manage the widest linguistic and factual knowledge directly applicable to their work, so that searching for information during the translation process is reduced to a minimum. TM systems naturally facilitate the integration of lexical resources into the system, by allowing the user to create archives (or databases) of past translations, terminology lists and lexica. Because lexical resources are so important to translators, systems offer various ways to build and access resources.

According to TM Survey 2006 [Lagoudaki, 2006:22], the significant majority of surveyed TM users (74%) stated that they prefer to fill their TM database, which is initially empty, gradually as they translate (i.e. by saving each segment they translate to the database automatically). 51% reported aligning legacy translations with their originals in order to produce a translation memory, 44% frequently import their client's resources and 28% import their company's resources. Downloading pairs of parallel texts from the web and then aligning them appeared to be an option for 15% of the users, aligning a ready-made bilingual parallel corpus was chosen by 12% and 6% used an external TM database that existed on a remote server.

The surveyed translators appeared to enjoy the *"st ructure and control [a TM system] provides over language resources"*, but many contemplated the fact that *"it takes a lot of time and effort to build data in a TM"*. In fact, the latter appeared to be the main reason for some translators for not wanting to use a TM system in their work. Overall, translators and terminologists alike showed great interest in being allowed by the system to build resources easily and quickly in many different ways. Specific needs regarding the most frequently mentioned ways of building resources – text alignment, terminology collection and term extraction – are discussed below.

5.1.2.1 Text alignment

It is commonly believed that existing translations contain better solutions to a wide range of translation problems than other available resources such as dictionaries, glossaries, and so on [Gough, 2002:1; Isabelle, 1993]. This, obviously, depends greatly on the size of the archive/database as well as the subject of included texts. Large repositories of legacy translations in subjects relevant to the area of expertise of the translator are normally expected to have immense potential for content recyclability, producing benefits such as time savings and quality assurance through consistency checks. Translators were found to prefer using their own legacy translations, if relevant to the text they are translating and if easily accessible, as they believe that this is a reliable source of information, of which they can be sure of the quality. When imported into or associated to a TM system, legacy translations are normally used for matching the new text against any previously translated similar text or for reference purposes only.

A significant number of TM users (51% of surveyed TM users) nowadays rely heavily on the alignment module of TM systems in order to build bi-textual databases or archives. Based on the survey's findings, the expectation and general need of all surveyed TM users regarding the alignment facility is the automatic accurate alignment of pairs of texts (in any file format) at all levels (e.g. document, section, paragraph, sentence, phrase, word) or at any particular level specified by the user. Interestingly, alignment is found to be one of few TM functions for which users favour automation. This seems to derive from negative experiences with current TM systems which assist the alignment process by offering alignment "suggestions" based on well-chosen parameters, but require sub stantial user intervention in order to verify and often repair the suggested aligned pairs. The parameters are typically punctuation and paragraph markers, repetitions, and non-linguistic matches such as numbers and abbreviations. Often, these parameters can be configured by the user. Still, alignment results are often disappointing (*"segments are divided in the wrong places"*), as several of the surveyed TM users have reported. Typical cases where manual changes are required are differences in sentence delimitation (one sentence in the source becomes several in the target or the other way around), shifts in the order of segments, different use and/or placement of footnotes, and index markers. Because of these cases,

many users have asked for the option to define segmentation rules, as well as more flexibility in merging, splitting and deleting segments (and undoing any of these actions).

Correct alignment is very important for correct bilingual match retrieval (see section 1.3.4.5). Yet, the process of alignment is considered very resource-intensive (time and effort consuming) and tedious, especially for lengthy documents or a great volume of past translations. Increasing the speed of the process was one of the main requests from TM users regarding alignment. This is obviously important as they expect a TM system to begin to deliver benefits immediately. Specifically, users believe that time can be saved if the system:

- can identify the language of texts automatically;
- can process unlimited pairs of files at once;
- employs a sophisticated alignment engine (performing linguistic analysis) that produces high accuracy of alignment, saving the need for manual alignment validation;
- filters out any unlikely match (for instance, based on differing lengths of segments).

The second most compelling need of the surveyed TM users is found to be the retention of the context for aligned segments. Context seems to be an ambiguous concept that is interpreted differently by various TM developers and linguists. Many TM developers presume that context refers to the extra-linguistic information that the system or user can attach to aligned segments (e.g. client, target readers, subject, date, other notes). However, linguists generally concur that context is not limited to the extra-linguistic environment in which communication is situated, which involves ideational knowledge of the world, but it also includes the co-text, i.e. the immediate verbal environment of a segment or term, which plays a crucial role in inferring the meaning of this segment or term [Hoye and Chan, 2007: 254; Baker, 1992:238]. TM developers certainly acknowledge the value of contextual information, but many appear to underestimate the importance of co-text. As a result, many of the existing TM systems today fail to present the aligned segments in their original textual surroundings, because the alignment has broken down the texts into segments, so the texts are nowhere to be found in full.

Translators seem to be suffering from this approach to alignment, as the retrieved matches devoid of their co-text can cause confusion. The context of a word or phrase, as Baker points out, determines the range of implicatures that may sensibly be derived from it [Baker, 1992:238]. If no context is retained, the translator cannot tell with certainty whether the disjointed chunks of texts that are presented to him as matches are appropriate for the text he is translating. Even if the segments are annotated with some basic contextual information (e.g. client, subject), the co-text is the defining descriptor of the

communication situation in which the segment or term occurs. Without the co-text, segments are just static specimens of a language, open to limitless interpretations, which may not help translators whose primary interest is not to examine the usage of a term or phrase – as lexicographers are often interested in instead – but to make pragmatically-informed decisions about the translation of the text in hand.

In fact, TM users who reported using systems that take a full-text approach to alignment (i.e. systems that do not separate sentences/terms from their co-text at the segmentation phase of alignment), seemed to appreciate the speed of building bi-textual archives/databases, and most importantly, the possibility of viewing the context for each suggested match, which helps them evaluate the system's proposals more quickly and with greater confidence.

In terms of storage options for TM resources, there seems to be no clear inclination, overall, in favour of a particular type of repository that may contain bi-textual resources (e.g. whether the TM application creates a database or simply indexes separate files that exist on one's hard drive without creating a database), as long as the application can handle very large resources and can demonstrate very fast retrieval times. Nonetheless, several complaints were expressed about the weak database structures used by some of the current TM systems, which lead users into frustrating situations such as dealing with corrupted databases or lost segments.

For those who favour the database solution, the kind of database implementation is assessed, firstly, from its ability to treat uniformly as many languages as possible and, secondly, from the ability to port to other systems seamlessly. The latter appears to be a compelling need for all types of translation professionals, who wish to be able to import and export resources into and from their system with ease, in a standard format, without having to convert proprietary database file formats to the standard ones. Project managers and translation company owners seem to be the ones who are more concerned about database issues, especially when they are faced with TM data migration scenarios (for instance, when they want to change to a new TM system). For many of the surveyed TM users, the solution to content storage appears to be the automatic conversion of all content to an XML format and its subsequent storage in an XML-aware repository.

An interesting idea that emerged from TM users with experience in using text retrieval software (e.g. Google Desktop Search, Copernic Desktop Search), was the creation of an index that contains only the text contents of one's bilingual resources, after those have been aligned at user-specified levels (e.g. paragraph, sentence), instead of the creation of a database containing disjointed aligned segments. Since the segments would not be separated from the full text, the system could show retrieved matches in full text (bi-textual scrolling would be a desirable added functionality), and what would be considered the most important advantage of all is the fast search and retrieval. The index could be updated

automatically to ensure that one's latest work is included, along with any updated glossaries, reference materials, even e-mails.

Finally, TM users appear to enjoy the flexibility of creating several separate resource repositories (for better organisation of resources or to facilitate sharing), as long as these can communicate with each other and can be operative simultaneously.

5.1.2.2 Terminology collection

Terminology research and management can take translators a considerable amount of time within a project[41], and can often turn into the most challenging and time-consuming task prior to or during translation, especially when translators are novices or when they undertake a translation in an unfamiliar subject area or for a new client. This is why any assistance on the terminological aspect of a translation job is greatly appreciated by the majority of translators and terminologists alike.

One of the most important utilities of a TM system is that it can act as a repository for consolidating and storing terminological information for use in future translation projects. The terminology module in every TM system has three basic functions: it allows its users to a) create and maintain terminological records, b) search for terminology before and during the translation process, c) import and insert terms automatically, eliminating the need to copy and paste or re-type terms, and d) export terminological data as dictionaries, glossaries or wordlists.

The most compelling need expressed by TM users concerning the terminology management function was the ability of the tool to process terminology in such a way that it can be made readily available and easily accessible. In other words, users need to be able to build termbases fast and with ease. For some, this means being able to manually select corresponding terms, by perhaps highlighting terms in both source and target texts, and using a shortcut or a mouse click to enter glossary terms in their terminology database. This could be done both as one works on the translation and after one has completed the translation. For others, it means being able to copy and paste terms from other applications (e.g. from a Word document, an e-mail or a webpage) into the glossary, without leaving the text one is working on. The ability to store and correct terminology on-the-fly is very important for the timely maintenance of terminological records. Obviously, for all TM users, the glossary import function in every TM tool is considered of paramount importance as it is the fastest way to build substantial terminological resources that can be accessible from within the TM system.

Other frequently-mentioned requests from TM users included the possibility of viewing all terms in a glossary at any point during translation and the option to reverse bilingual

[41] t-survey 2005 found that translators spend on average approximately 30% of their time on terminology work (investigation, storing, validation, etc.) in relation to the duration of the translation process [Zielinski & Ramírez Safar, 2005].

glossaries (for example, a Spanish to English glossary could be also used or exported as an English to Spanish one). Those who often have to type terms in their terminology databases also found the automatic change of input language once the cursor is in the translation field to be very useful (saving the user from manually changing between languages each time he wants to key in the translation for a term).

The second most compelling need expressed by TM users was having flexibility in the type and amount of information that can be stored on a term record. Although terms are generally considered as monosemous – and thus more likely to have one translation, different customers may use different terms to express the same concept. Moreover, glossaries do not always contain industry-specific terminology only. As Esselink points out, project glossaries can sometimes also contain product names that should not be translated, words and even phrases that are repeated throughout the project and equivalents for general language verbs and nouns that are appropriate for the product [Esselink, 2000: 402]. Mapping source-to-target terminology using a unidirectional 1:1 correspondence is therefore not always practical, as a term or other word/phrase found in a glossary may have a number of definitions and translations depending on context. On this basis, a concept-based terminology bank seems to be the preferred solution.

Furthermore, users who find themselves limited in the type of information they can attach to each term (by filing in predefined set of fields, such as subject field, source, definition) have reported that such a limitation impacts on the reliability of their terminological resources, as the less information is presented about a term, the less confident the translator feels when trying to decide whether to use it or not. On the other hand, configurable glossary structures which allow users to define their own fields of information, seem to assist translators organise their resources in a more efficient manner. In particular, TM users find very useful the ability to add grammatical, semantic, syntactic and pragmatic information to their terms, and even link them to external reference sources (e.g. Wikipedia pages). Concordance statistics offered automatically by the system for each term appear also to be of help during decision-making. Some also find very helpful any system-added sourcing info for each term (e.g. file name, folder location).

Finally, a significant number of TM users appeared to favour and consider of extreme value the solution of a temporary terminology container (sometimes called a 'lexicon') that some TM systems offer. This type of container acts as a temporary holding place for data that still needs to be verified before being sent to the terminology database, or as means for extracting terminology before starting a project (the system creates a wordlist of all words/phrases which occur in the text one is about to translate), or simply as a means to import data quickly into the translation memory or terminology database. This provisional wordlist is particularly helpful to those who want to have a glossary with the frequently-found words/terms for this project straight after they complete a translation.

5.1.2.3 Terminology extraction

Enlargement or improvement of one's available dictionaries and glossaries can be helped by terminology extraction from parallel corpora or one's previous translations. Terms that are extracted from approved translations normally inspire confidence about their quality, compared to glossaries obtained from unknown sources; they can therefore help translators and terminologists control the quality of their terminological resources. Term detection and extraction is a labour-intensive activity which can be assisted by the terminology extraction tools often incorporated into a TM system. Such tools facilitate the time-consuming process of terminology compilation, by scanning lengthy textual resources, detecting terms and their corresponding translation and presenting candidate terms to the translator or terminologist for validation. The methods that each tool uses vary between statistical techniques and those which make use of linguistic knowledge (e.g. morpho-syntactic and semantic resources). Hybrid systems may also combine techniques from both approaches, while some other systems have appeared to be using exogenous methods, such as a contrast corpus for identifying terms in a specific domain [Oliver and Vàzquez, 2007: 6-8].

The surveyed users appeared dissatisfied with statistical approaches to terminology extraction, as applied by current TM systems. Those offer an extraction feature for frequent terms, for glossary building purposes:

> *"[I'd like to see] better 'terminology extraction', i.e. extraction not based only on number of occurrences of a term in a text, but 'technicity' of the term (which may appear only once or twice but will require a lot of research to find the appropriate equivalent in the target language). The existing tools right now only make up lists of 'frequently found' terms, which 99% of the time are not technical terms related to a specific technical topic."*

On the other hand, linguistic approaches that make use of linguistic knowledge (or combined linguistic and statistical methods) are thought to produce better results. Ideally, users would prefer an accurate term extraction, where all and only the relevant terms are extracted from lengthy texts at great speed. But to date no terminology extraction method has managed to dispense with human input altogether, human intervention being present in all terminology extraction applications to a greater or lesser extent in order for the system to produce the desired quality of terminology. This is natural since the type of information to be extracted depends each time on the areas of specific interest to each client or user.

Thurmair [2003] observes that the relevancy of candidate terms (proposed by a term extractor) is a highly idiosyncratic issue. This is not only because terms are selected and stored on the basis of the likelihood of their being used in the future, but also because when translators compile a glossary with terminology, they not only include terms (from a special domain) but they also often include general purpose language words that they do not know, or had difficulty translating. So the term extraction task will always be linked to the particular translator's knowledge as well [ibid, 2003].

However, the survey found that terminology extraction is one of most desirable functions in a TM system, especially for translators, terminologists and managers of translation departments within large organisations. This finding confirms similar findings from the t-survey 2005 on terminology extraction practices and tools conducted by Zielinski and Ramírez Safar. In their survey, although only 15% of the 451 respondents reported using a terminology extraction tool for their daily work, 85% of all respondents expressed potential interest in using such tools as they recognised their potential benefits [Zielinski & Ramírez Safar, 2005]. This disparity in figures might seem remarkable, but it is not surprising when one looks at the reasons that non-users gave for not wanting to use any of the existing terminology extraction tools. The most common reason of all was the tools' poor performance in extracting terms. In particular, respondents mentioned that the results lists of the terminology extraction usually contained too many wrong term candidates, or, in other instances, the tools found very few terms. In addition, the time spent checking the list of term candidates is often too long to afford on a regular basis. Finally, the terminology extraction results list often did not contain adequate information about the terms, which would allow the translator or terminologist to decide with certainty whether a word was a term or not.

According to the same research, desired functionalities for a term extractor included (in order of importance):

- linking the term candidates directly with their contexts: this appears to be of paramount importance according to our survey as well, as terms can be ambiguous at times (many terms are often used metaphorically in general purpose language, or the same term can be used with different meaning in different domains); our survey confirmed a vital need for viewing the context of terms (if extracted from texts) indicated by the great number of respondents who reported performing term concordance search on a very frequent basis because the context of a term was not shown;
- searching automatically for definitions of terms (e.g. with the help of internet search engines);
- communicating with the termbase in order not to propose already known terms as term candidates;
- showing collocations of the term candidates;
- proposing synonyms and antonyms;
- automatically identifying linguistic information;
- showing semantic relations for the terms contained in the termbase or in the text;
- displaying a notional system as a diagram in addition to the results list.

In addition to the above, our TM survey's respondents expressed interest in additional functionality such as:

- selective extraction of terms (e.g. noun phrases or verb phrases only);
- recognition of idiomatic expressions;
- recognition of jargon (*"recognise the difference between 'dictionary' words and management rhubarb/invented words - very predominant in German"*);
- recognition of special entities such as names of companies or people, numerical data, dates and places;
- automatic coding of terms (e.g. if a corpus comes from a specific client or subject, the extracted terms should be automatically tagged with such information to facilitate future searching).

In essence, a terminology extractor needs to be based on sophisticated natural language processing components that utilise language resources specific to the working languages and domains of the TM user. It needs to be trainable and intelligent enough to extract not only terms from monolingual texts but also term pairs from bilingual comparable or parallel corpora which can be easily built from the web. Recent attempts to extract pairs of terms from bilingual corpora [see Ha et al., 2008a:107-112; Ha et al., 2008b], as well as from comparable corpora [see Saralegi et al., 2008: 273-280; Gamallo, 2007:191-197] have presented some very encouraging results to this direction.

5.1.3 Resource management

Managing and controlling the quality of language resources residing in one's TM repository (e.g. glossaries, aligned past translations, reference material) is vitally important as, without the appropriate maintenance, there is a real risk of quality deteriorating, especially if the repository is fed by a number of translators. This is often the case with translation companies or departments which use a number of translation vendors or suppliers. In many cases, also, the client wishes to provide his own translation memory which, if it contains errors, can contaminate the translator's resources. The consequence of poor resource maintenance in all these cases is either that productivity is impaired by inappropriate or incorrect matches or, worse still, that less experienced or inattentive translators adopt inappropriate translations.

The surveyed TM users made a number of suggestions for improving the TM maintenance features, stressing the need to be able to manage their resources in an easy and effective way.

5.1.3.1 TM maintenance

A repeated complaint by the users of TM systems is that the memory does not get updated properly during translation or translation editing/proofreading. Often, the final changes to a translation do not make it into the memory, or do not overwrite old incorrect translations that exist in the memory (being instead stored as duplicate translation units). Accessing and maintaining the data of the memory directly from the main interface appears to be very important for TM users, as it facilitates applying changes in the TM when translating or reviewing the translated text. They also seem to appreciate the ability to batch edit their memories interactively, directly from their translation environment.

In essence, the TM maintenance module should be able to show source text, target text and any attributes of the translation unit (with the ability to hide the latter if one wants to keep things simple) and should allow the user to scroll through all the translation units and edit them all as desired. The module should offer also the ability to select translation units that are going to be batch edited automatically, as well as a Find and Replace function, to increase the speed of 'cleaning' large memories. Being able to merge two translation units as one or split one entry into as many segments as one wants is necessary for obvious reasons.

The module could also flag translation units according to a variety of conditions which are frequently encountered (e.g. flag entries with the same source, target or both (duplicate entries), entries with empty source or target, entries where the target is the same as the source). Flagging duplicate entries is particularly helpful for the translator, as he can check whether there are incorrect translations in his memory that should be deleted or whether there are alternative translations for the same source that should be kept and used in different contexts or types of subject matter. Once these translation units have been flagged, the user could filter them so that only they are displayed, making it easier to correct them (or delete them all at once, if required). The user must be able to specify his own flagging conditions, including with relation to client and subject area fields if present.

5.1.3.2 Terminology maintenance

Like the translation memory, terminology is easier to maintain when terms are accessed from the translation environment and edited in synchrony with the task of translating or reviewing:

> *"Terminology-management tasks [should be] performed conversationally within the translation environment (being able to edit [name of terminology management tool] from the [name of TM system] environment so as to handle multiple equivalents for a source term, e.g., instead of editing the entries ex post facto)."*

A dedicated terminology maintenance module should allow the user, apart from editing the terms, to add and remove glossary fields, such as definitions, client or subject, and even

blacklist terms that are of poor or dubious quality. Users also expressed an interest in being able to specify the order in which the programme lists suggested terms (i.e. assign an ordinal rather than only having the choice of chronological order or reverse chronological order). User could even decide which fields of a term record are visible in the terminology window in the translation environment when a term is found in the text, e.g. context, translation comment, definition.

To ensure terminology organisation and control, the module could allow the user to filter all terms in a glossary by any combination of user-specified criteria (e.g. by client, subject, translator, date of entry creation) and flag duplicates, terminology inconsistencies and blacklisted terms. In addition, other useful capabilities as requested by the surveyed TM users may include:

- allowing multiple termbases to be merged or split according to criteria,
- generating reports and metrics on any variety of terminology-related matters,
- generating a diagnosis of the status of the glossary
- creating dynamic links between resources:
  - concordancing selected terms against the translation memory and locating examples of use in order to check the validity of their translation,
  - comparing two glossaries for missing or inconsistently translated terms,
  - performing terminology synchronisation between two glossaries,
  - displaying the frequency in which terms occur in all terminological resources,
  - creating cross-references between terms, or between terms and specific texts in a memory.

5.1.3.3 Organisation of resources & sharing

A persistent requirement by TM users is the ability to create translation memories or terminological resources separately from any particular project (i.e. not having to create 'project kits'). The reason for this is to facilitate the sharing of resources in neat packages. To improve organisation, the resource management module should allow the attribution of custom labels for individual resources (e.g. translator, project number, client) and filtering can be applied to select and batch edit the resources that the user desires. For more effective batch editing, different filtering criteria could be combined. In addition, users must to be able to merge or split resources by filtering their entries. This seems to be particularly useful for project managers and clients who often wish to supply extracts of their translation memory files to freelancers.

Finally, the module must allow the comparison of resources in the case where different translators have been working on and simultaneously updating an offline TM resource file (e.g. a TMX file) and need to merge it back into the original translation memory after they finish their translation. This way, translation managers can check and update their translation memories more easily and on a more regular basis.

The resource management module seems also to be helpful of improving the quality of retrieved matches, if either or both of the following two facilities are offered:

- the user is able to sort resources hierarchically according to his criteria (this helps weighting the importance of matches when retrieved and suggested to the translator);
- the system generates reso urce rankings according to the frequency of use of a resource (displaying the most frequently used resources at the top, and prioritising their matches accordingly).

The most important need of all in relation to resource management is the ability to import and export resources in multiple formats - e.g. XML, plain text or HTML - for the user's own use, but also in whatever TM or database format the clients might supply or demand. Adhering to open resource exchange standards is perceived as the best solution insofar as it makes the exchange of resources and integration within other applications easier. For the more demanding users (who work mostly on collaborative projects), a TM system must also offer the option of publishing resources over a server to be accessible in real time by any standard internet browser.

5.1.4 Search

The search facility is key in every TM system as its efficiency impacts on the effective retrieval of matches. The surveyed TM users offered suggestions as to both *where* and *how* the system should be able to look for matches.

A frequently mentioned request was the possibility for the user to mix and match the resources that he wants to associate with a translation project without any limitation on the number of resources selected. The user should also be able to rank the selected resources, so that matches from top-ranked resources are displayed first. An interesting idea that some users seemed to favour was the ability to distinguish and select 'primary' resources (for matches that can be trusted) and 'secondary' or 'reference' resources that may consist of not so trustworthy legacy translations or glossaries, or of monolingual or bilingual corpora, or even external resources (e.g. glossaries or texts on the web or a remote server) provided that those can be accessed in some way. The ideal scenario for the surveyed translators seems to be the simultaneous searching of multiple resources:

> *"I would very much like to be able to centralize the content of all the dictionaries I have purchased and of glossaries available free on Internet, in order to query against them as I translate and have the possibility to insert translation of terms in my TM environment."*

Translation managers and some translators alike (particularly those working in collaboration with others) favour also the option of 'locking' resources, so that they cannot be accessed in cases such as when being suspect, duplicated or company-specific. A visual component showing which resources are 'live' and which ones are 'locked', along with short property descriptions is important for the translator in order to maintain control of his sources of information.

Two types of searching should be possible: a) for matches and b) for context (parallel concordance search). An option must be made available to operate either of the two or to set the concordance search to start once matches have been found (especially for those matches that do not have any co-text or context). In addition, the user must be able to select either an automatic or a manual search for either of the two types of search. This appears to be very practical for translators or reviewers who wish to use a TM tool for concordance searches only or those who prefer more control, by searching only the unknown terms/phrases and not everything. It is also important that the search field(s) are detachable from the main interface, and that the search module can be used as a separate mini-application for when someone translates from hard copies or wants to translate in other environments.

Searching for matches is normally performed automatically in virtually all TM systems (e.g. when a term/phrase is highlighted or a segment is selected), which is why the speed of searching is one of the primary concerns among TM users, especially for those who own very large resource repositories. The flexibility of searching is also one of the top requests, with users favouring the idea of having various different look-up methods to choose from and combine. From their responses to the survey, at least three search methods should be possible:

1. Simple search: for locating specific words/phrases, even when those exist within larger textual structures, like sentences, in the translation memory. Strategies employed by the TM system should include sub-segment searching, meaning that every word or combination of words in the source text can theoretically be searched for. The system should be able to match text at any level: word, expression, sentence and paragraph in full-text context:

> *"I think a good TM tool should do matching at 5 levels: 1. Paragraph to Paragraph TM 2. Sentence to Sentence TM. 3. Phrase to Prase TM. 4. Word to Word TM. 5. Character to Character TM. If 1 is not good match, then it should look for 2. If 2 is not good match, then it should look for 3. If 3 is not good match, then look for 4. If 4 is not good match, then look for 5. "*

2. Fuzzy search: for locating words/phrases that are similar or contain part of the search text. An advanced fuzzy searching technique will have linguistic capabilites (e.g. recognizing singular-plural forms, inflections, etc.) that will help to overcome some of the challenges in searching within specific languages. Every language has its own set of special challenges when it comes to search retrieval. For instance, as Halpern [2004: 60-1] explains, Japanese has highly irregular orthography, so orthographic disambiguation is essential for achieving high performance and precision in cross-language information retrieval. Cross-orthographic searching is a requirement that can improve the match recall rate dramatically. Depending on the fuzzy matching performance of the system, the translator must be able to set the sensitivity threshold of the match, so that a balanced number of potentially useful matches are retrieved.

3. Wildcard search: for locating all words that contain a specific string of characters. Wildcards are very useful for highly inflected languages. For example, a termbank may contain only one form of a term (e.g. 'exctracting'), while the same term in a different form (e.g. 'extraction' or 'extractor'), if looked up, may retrieve no matches. Using an operator as a wildcard after the root part of the term (e.g. 'extract*') will return all available terms having this root, which can also be helpful to the translator.

Other options for improving flexibility in match searching would be allowing the user to select whether the search string is case sensitive or not, to find whole words only, and also to specify the number of matches displayed for every search query.

In terms of parallel concordance search, it appeared to be a very popular feature with the respondents of this survey. In particular, users of TM systems that make use of translation memories containing disjointed segments (instead of full bi-texts) find the use of a concordance tool indispensable, as it can show some co-text for the available matches. The function of the TL word/ phrase in its specific context is important for the decision-making, whereas in the absence of contextual information, the suggestions can confuse the translator.

According to the surveyed TM users, the usefulness of the concordance search function can be improved if, apart from creating wordlists with statistical information on the term/phrase one is interested in, it can also:

- Show the occurrence of a given search term/phrase in its co-text (i.e. the words to its right and left) and allow the user to cutomise the length of text displayed before and after (with a "more" option to reveal the full text, if available). This way, the translator can search through all the selected resources at once to examine every occurrence of a particular term in order to understand how it is being used. This

saves him a lot of time and allows him to grasp the meaning of a complex concept immediately and decide upon its correct translation.

- Filter results by client, subject and so on (it being possible to combine different criteria), in both source and target languages, or filter on the results that are of interest (e.g. the relevant translations of the queried term). By filtering on the results of interest, the user can concentrate on these results only and batch edit them if necessary, without running the risk of inadvertently changing any other translation units in the project.

In addition, a number of search options must be given to users, so that they can improve the quality of their search results:

- use of wildcards for finding inflected forms of a word
- choice of ignoring or matching punctuation marks (punctuation marks are normally ignored when matching, but in some cases, e.g. for capital letter acronyms, punctuation marks must be matched as well, in order to retrieve the correct matches)
- choice of matching upper-case or lower-case letters or both
- choice of method (e.g. Boolean operators) for combining adjacent search terms.

5.1.5 Translation assembly[42]

More and more TM systems nowadays are fortified with machine translation techniques to enable them to propose a translation to the translator when no match is found in his TM resources. The system attempts this by assembling a combination of terms from its terminology database, translations from its memory, and even portions of them, with the help of integrated MT techniques.

Translators are different from the average user of MT systems when it comes to evaluating the machine translation capabilities of a TM system, therefore they have slightly different requirements. Generally, the latter are just looking for the gist of a text, or want a quick translation of a short-lived document. By contrast, translators are expected to produce an excellent quality translation on every occasion. They can use MT if they like, but the final result should be seamless. This means that the standard of editing is high, and if the MT output is poor, correcting it might take as long – or longer – as starting from scratch. So, while the average user's MT tolerance threshold is normally moderate, the translator's

[42] This section has been published in a modified form in the proceedings of The Eighth Conference of the Association for Machine Translation in the Americas, Wakiki, Hawaii, 21-25 October 2008. Paper title: "The Value of Machine Translation for the Professional Translator" by Elina Lagoudaki.

threshold is much greater, and what is useful for the one may not always be helpful to the other.

An important issue that impacts on the MT capabilities of a TM system, in contrast to that of standalone MT systems, is that the performance of the system stems from a complex interaction between the processor and the input, i.e. the translator's own resources. The success of the MT functionality, therefore, depends considerably, among other things, on the size and quality of lexical resources residing in the TM system. On these grounds, the surveyed TM users were not asked to rate the accuracy of match construction as performed by their TM system, but instead to offer comments and suggestions on the practical usefulness of MT functionality as part of a TM system.

Generally speaking, machine translation appeared to be well received amongst translators who were familiar with it, with 45 of 145 TM users (using a TM system with MT capabilities) declaring it to be their favourite feature in their TM. For these users, assembling sub-segment portions from their terminology database as well as their translation memory (thus rendering it possible to have a 'match' even if the exact source text is not in the TM) is an advantage that increases their productivity and saves them time. Even when the results of the assembly are not 100% accurate, some of these users are still satisfied, as access to a couple of translations – even if they do not make sense – can bring together a number of useful clues which they can utilise before they select everything, delete it and type in their new translation.

However, there was also a small proportion of respondents (10) who named the automatic assembly feature as the most annoying of all. These users were mainly dissatisfied with the repeated poor quality of match assembly. As one user observed

> *"...of the many previous translations for a given expression, some will fit in the current context. However, many examples that apply in specific contexts will also exist but they will not be fit because the terminology usage, style, and/or tone are inappropriate."*

Another user, along similar lines of dissatisfaction, expressed his concern about the lack of linguistic and stylistic unity in a translation assembled from different pieces of text, something that does not only happen to literary translations, but also to non-literary ones. Indeed, most TM resources are made up of many different texts that might have been produced by many different authors and translators. Each text and each translator can have a different style, and when bits and pieces from a variety of texts are forcibly brought together, the resulting text can be a stylistic patchwork.

The attitudes of the respondents who commented on the MT features of a TM system varied, as expected. Some appeared to dismiss the whole idea of machine translation functionality in a TM system:

> *"I don't think that a TM tool is supposed 'to bake a cake' while I am translating. There is a reason why it is called a translation memory tool and not machine*

> *translation. I believe that most of the things mentioned above [MT functions] are supposed to be handled by human translators. Sometimes less can be more."*

On the other hand, other respondents expressed the desire to see more advanced MT capabilities:

> *"A well designed translation program should not be based on the mere mechanical replacing of texts... but it must think, analyze and reconstruct the text in the destination language, using its database to enhance the translation."*

The respondents of the second category were those who offered recommendations as to how MT capabilities could be strengthened to serve their needs more effectively. Some examples are listed below:

- A machine translation facility should be available when there is no match or the fuzzy match is below a user-predefined percentage – say 75%. The machine translation should get the information from a separate database which will be full of parallel corpora. In the TM options the user would be able to choose two databases: the client or project-specific one (as primary) and the master one (as secondary) which will contain the parallel corpus.
- The TM tool should have linguistic capabilites (e.g. recognizing singular-plural forms, inflections, etc.). When, for example, it finds fuzzy matches from the TM, it should have the ability to analyse and do some operations on the target segment as well (e.g. automatically correct the grammar).
- In order to assemble semantically precise and relevant translations, there needs to be in place some measure of context similarity between the source phrase and the retrieved matches.
- Incorporate MT at a user-defined level for sub-segment/new segment processing (e.g. search for terms in standard dictionaries).
- Have the option to turn any MT functionality on or off according to one's preference.
- The system should identify similar examples and make generalisations about them.
- Integrate automated and user-dependent feedback of new knowledge into the knowledge base in a dynamic and interactive manner.

By correlating the attitudes of translators (positive/negative) with various factors that may affect the way they view machine translation, some interesting findings emerged. There was no evidence to suggest that a positive or negative attitude is related to the age or the computer usage competence of the translator. However, attitudes appeared to be affected by the length of work experience of translators. Inexperienced translators seemed to favour machine translation and to tolerate incorrect assembly more habitually. Experienced translators, on the other hand, with more than five years of translation experience, were

those who voiced complaints and who seemed to be most annoyed with machine translation suggestions. This group of translators generally thinks that it is mundane to wall on someone else's (the system's in this case) path. It is much more gratifying to come up with a decision on their own. They insist that translation requires analytic and intuitive decision making, and they prefer to make a decision on the spot, execute it and move on, without spending time in going through poor, or irrelevant or misleading potential translations.

In addition, attitudes towards machine translation seemed to be related to the language combination a translator worked with. For example, translators working with highly inflected languages, such as Greek or Polish, realised that machine translation underperformed (and consequently was of little use), unless it could use language-specific assembly algorithms and rules. Their expectation was that when a TM system incorporates a knowledge base (grammar rules, lexicon) it will produce more lexically and grammatically correct translations.

Finally and not surprisingly, the users' attitude towards machine translation appeared to be influenced by the area of specialisation of the translator. Those who regularly translated highly specialised texts or updates of previously translated texts seemed to be positively inclined towards the use of MT features and they would be interested in training their system, by supplying feedback, in order to improve its performance.

As appears from the survey's findings, the majority of translators hardly ever need functional machine translations (that may allow them to get the gist of a text), as they prefer to resort to authoritative dictionaries and other sources of pragmatic information for that. Instead, they expect from a machine translation feature the ability to suggest correct translations: to construct not only a correct segment syntactically and grammatically, but also in terms of semantic equivalence to the source text. Then, the translator will be the one to decide whether the suggested translation is fit for context by looking at the full text in hand. Anything less than a correct translation can be misleading or, if it happens repeatedly, can reduce the role of the translator to that of a post-editor of badly translated texts.

It is also evident from the users' recommendations that it is important for the system to be able to learn from the decisions/choices made by users: for example, which potential translations are preferred, which were rejected and why, with the system updating the weight for the selections made by the user, so that errors in future translation assemblies are reduced.

Although when MT systems first appeared they provoked negative feelings on the part of suspicious translators, the latter have now started to realise the benefits of using MT features, as long as *they are in control* of a) the repository content (thus ensuring reliability

of results) and b) the operation (by defining the rules of match construction and even opting out of MT matching when it is not necessary).

Translators also seem to be coming to terms with machine translation as an alternative means of translation production and appear to feel secure in their profession, reasoning that a fully automatic MT system is unlikely to produce high-quality translation of a text in the near future. This derives from their belief that even if computer programming succeeds in addressing even the most complex equivalences with appropriate descriptions of all possible discourse environments/situations, it can never accurately predict the aims, purposes, intentions, complex strategies, changing tactics and fine choices required by the creative character of a high-quality translation. Furthermore, translators know better than anyone else that the linguistic choices in a translation are a function not only of language rules but also of the translator's creativity, a factor that knows no rules and therefore cannot be replicated.

However, since not all texts involve elements of creative writing (many are rather repetitive, tedious and written in controlled language), translators have no reason not to appreciate the assistance of machine translation in dealing quickly with such texts, which offer little joy of intellectual challenge in any case.

5.1.6 Project management

Translation projects can range from a two-page letter in MS Word format to a complex multi-media website which includes a variety of file formats containing translatable content. As the complexity of a project increases, the translator, or the project manager in a translation agency, feels more strongly the need to be able to manage his project in a streamlined way, having access to a variety of practical tools provided by his TM system.

5.1.6.1 Project setup

The most compelling need on the part of the surveyed translators and project managers alike is the ability of their TM system to open all types of file easily (and if the project consists of many different files, to open all within the same project), allowing the user to work on them immediately, without having to spend time on file conversions. Loading more than one file in a project is also very useful for any project where the client sends several files, allowing the translator to search all the files at once for every occurrence of a particular term in order to understand how it is used. Global changes should also be possible in a project containing multiple files (e.g. finding a translated word in several documents – in the particular project – and replacing it with a new translation in all documents at the same time).

All elements within a project should be variable and substitutable, such as changing the languages in a project, with the possibility of having unlimited source and target languages in the case of multilingual source documents, or changing the number and types of resources associated with the project, at any time. The system should also allow the user to add, remove or update files in his projects as they progress. This kind of flexibility can be very useful when translators are working for clients who keep revising the documents they have sent as work progresses. This is not at all unusual in industries where time to market is crucial.

A TM user should also be able to attach localisation directives, notes or an in-house stylesheet to a project. Some organisations have in place a systematic documentation management system that includes the use of a stylesheet for authors and translators to follow for all documentation. This ensures consistency in the use of terminology, and adherence to pre-determined phraseology ensures uniformity of style.

A further request expressed by a number of translators and project managers was the possibility of keeping the files ordered in folders and subfolders within a project, thus making it easier to organise large and complex projects. The sub-division of the entire project into smaller projects would be very practical for distributed translation projects.

In order to monitor the progress of a project, several attributes could be attached to each file (e.g. wordcount, number of words translated, reviewed, signed off), allowing an overall view of how the project is advancing and its state at any time. A project status report can be generated and shared amongst a network of distributed translation teams and the client.

Analysing the project's source files for leverage is an important step at the project creation stage, as it helps the translator, or project manager, understand what savings can be made from the use of a TM. An integrated analysis function should be able to identify the translatable from the non-translatable parts of a text and report on:

- wordcount for translatable text (for each file separately and for the entire project in total)
- external matches (full and fuzzy) with the associated resources (for each file separately and for the entire project in total)
- internal matches (how many segments are repeated within the same document and/or within the project)
- number of images contained in the project which may contain translatable text (translators have reported that they often forget to translate images embedded in a document as their TM system skips them as untranslatable material)

For freelance translators, being able to apply the TM and have an approximate word count before starting the translation is very helpful as they can book their schedule according to

the work they really have to do. Some would also like to see an estimate of the time that will be required (based on an average value that the translator can enter) to translate the source files that they have just been asked to translate.

In general, the project management module of a TM system should offer functionality that allows full control over all stages of the translation, but also presents a simple way of creating a new project.

5.1.6.2 File format support

From the survey respondents' comments on file formats, the language services industry appears to be facing an acute problem which TM developers must address without delay. In the current situation, an immense variety of formats is used to store data. Consequently, translators and translation agencies are confronted every day with format conversion. Most TM systems try to offer some converters (also known as filters) that handle popular formats such as PageMaker, FrameMaker, Quark, and PowerPoint, but no known system covers all formats, as a) there are too many formats (every application can have its own proprietory formats) and b) filters are expensive to develop. Even when filters are developed for the popular formats, two serious problems are bound to emerge for the TM user. Firstly, every time an application releases a new version, its formats evolve, therefore the filters for these need to be updated too (adding considerably to the cost of maintaining an up-to-date TM system). Secondly, filters are not perfect. Some sort of information is likely to be lost during the conversion.

Some of the survey respondents suggested that using XML as a gateway format to convert to/from any other format can be the solution, since many applications nowadays can export their files into an XML-based format which is translatable.

The most popular file formats for which the survey's repondents requested support by a TM system are listed below, in order of frequency of mention:

- plain text files (e.g. .txt)
- MS Office formats (Word, Excel, PowerPoint, Visio, etc.)
- editable PDF files (with the ability to export translated documents in the PDF format, which seems to becoming a standard file format for document exchange, as it is searchable, viewable and printable on any platform and can maintain information integrity)
- HTML, SGML, XML (which can have embedded HTML or Javascript)
- TMX, TBX, XLIFF files
- Quark, InDesign, PageMaker, FrameMaker files
- OpenOffice and Star Office
- editable image files (e.g. .psd, .svg)
- online help file formats (e.g. HTMLHelp, JavaHelp)
- software text-only resource files (e.g .rc, .properties)

- subtitling files
- in-game files that contain text assets
- .po and .pot files (translatable language resource files used in the free GNU gettext concept for translating software and documentation)
- AUTOCAD formats

The point that most translators made very clearly is that a TM tool should be able to translate the file formats supported without requiring the corresponding application to be installed. For example, some tools claim support for MS Office but cannot do anything if MS Office is not installed. They use MS Office as an intermediate filter.

Non-editable graphics

A challenge for every translator is presented when dealing with text that comes in a graphics file type (e.g. .jpg, .gif, .bmp, tiff) or a non-editable PDF. Since a TM system cannot process the text in these files (unless it can successfully use an integrated OCR functionality), it could allow the TM user to browse folders and subfolders in his project quickly for graphics that are displayed within the tool, determine which contain translatable content, create a list of those image files, and (manually) enter the source text. This can then be exported into a file that he can translate in something like Excel and the file can then be delivered to the client or passed on to a DTP specialist to localise the graphic.

The ability of a system to open this kind of file can bring a considerable advantage to the translator. It will allow him to use images as references when translating, either when he needs to translate them or when he needs to refer to them as a resource. In fact, many translators have reported that a part of their resources consists of scanned documents, non-editable PDFs and even scanned dictionaries, which could be used more effectively if they could be browsed from within their TM system.

5.1.6.3 Version control

It is increasingly acknowledged that with the rise of digital 'disposable' content, the translation process has passed from a linear to an incremental just-in-time model for a large number of translation projects [Désilets et al., 2006]. Material is now being translated before the source material is complete or finalised, so that the translation makes it for simultaneous shipment. Websites, whose information relies on its freshness, are a prime example of this new process, as they go through regular updates and accordingly language support needs to incorporate all such changes. This can become complex in terms of management and workflow, particularly when multiple languages are involved.

According to the surveyed users' opinion, a TM system should be able to provide assistance in update management and version control with a set of functions that are integrated in the project management module. Such functions would allow:

- the accurate comparison of two or more versions of the same source file, so that the translator can know which text has been edited and which is the new text to translate, with the ability to isolate the new text
- the identification, marking and analysis of overlapping content using a visual comparison utility.

5.1.7 Quality control

The quality of the end product of translation is of paramount importance to all translators and maintaining a high level of quality is in the interests of every translation company. Virtually all TM systems nowadays come equipped with some type of quality assurance (QA) tools. According to a survey by Makoushina [2007], QA tools are indispensable for the majority of modern translation professionals (81% of the respondents of that survey reported using QA tools), and 83% of those use the QA tools that are built into their TM system (as opposed to standalone QA applications). The same survey showed that 32% reported they were applying quality assurance procedures at the end of each translation stage (e.g. translation, editing, proofreading), whereas the second most popular approach (28%) was to apply QA procedures only at the final stage before delivering the files to the client. 30% of professionals and translation companies reported that they applied QA procedures to source files as well as to the final ones.

The last point was also highlighted by the respondents in our survey who recognise that poor language in the source text is often the reason why no matches are retrieved from the TM repositories. Poor source quality is also believed to impact on the quality of resources when source and target texts are subsequently aligned and stored to be used as reference. The same respondents are convinced that improving the source text will in turn increase the likelihood of better output in other languages and derivative works.

Given that translation quality is directly linked, to a smaller or greater degree depending on the skill of the translator, with the quality of the source text(s), on one hand, and the quality of resources used, on the other, to assure the optimum quality for the end product, a TM system needs to perform quality checks on all three levels: source text, target text and resources. In particular, quality checking and fixing external reference files before using them for a translation project and possibly multiplying the problems throughout the project appears to be a primary need on the part of translators, who are naturally concerned about the reliability of their resources.

The surveyed TM users suggested various options for running QA checks. Some need to be able to perform a QA check of the source document as soon as they set up the project, before they start any translation, for problem areas such as incorrect spellings or in-sentence line breaks. Others find the on-the-fly underlining of language errors in the source and target texts to be practical, as well as a real-time check of translated terms against the glossaries, while they translate. They also favour a customisable 'auto correct' or 'auto text' feature. Alternatively, other TM users find it more convenient to perform QA checks after they finish the translation of a document as part of their checking procedure, or finally as a batch operation on all the files in their project. The option that was less favoured by the surveyed TM users was that of QA checking every segment before it gets sent to the memory. Users found this option very impractical, annoying and time consuming, mainly because it constantly interrupted the task of translation.

Regarding the type of errors that a TM system should be able to detect and correct automatically, spelling mistakes were the ones most frequently mentioned by translators, reviewers and proofreaders. More specifically, these TM users expect the system to be able to run context-sensitive language checks with the possibility of learning from the corrections made by the users. A text-induced spelling correction method that uses nothing but lexical and word co-occurrence information derived from the translator's own collection of texts and translation memory, in order to perform context-sensitive spelling error correction of non-words, has been proposed by Reynaert [2005] and seems to come close to this particular need on the part of TM users.

Respondents to our survey have reported various additional translation errors whose detection could be automated by a TM system. These are listed in the table below (Table 10). The list is by no means exhaustive and only includes the errors that the survey respondents mentioned as the most frequently encountered.

Common error typology		
Language/locale checks	Consistency checks	Visual checks – formatting integrity
spelling errors grammar/style errors language-specific quotation marks language-specific punctuation spaces before full-stops non-breaking spaces (between number and measuring unit, for example) non-breaking space before : ; ? ! multiple spaces (except after...) no multiple dots (ignore ellipsis dots (...)) double punctuation capitalisation of initial letters measurement unit conversion	omitted translation incomplete translation insertion of translation that lacks support from the source text/ double translation identical numbers, figures, email addresses, names identical hyperlinks in source/target identical untranslatables in source/target identical punctuation at the end of source/target segment project glossaries/resources adherence identical spacing between titles, headings and paragraphs	errors in tags consistent font type, style and size correct placement and size of graphics page flow and page numbering cross-references between text and the table of contents, indices, internal references, graphic caption text text indentation and alignment correct display of fonts and graphics within appropriate browser or reader corrupted characters text direction display of copyright and trademark symbols all translator or reviewer comments removed

Table 10: Common errors detectable by a TM system

From the above list, checking for terminology consistency is also thought to be one of the most essential QA checks. Quality control can ensure that the translator or reviewer is alerted to any inconsistencies in terminology with respect, for example, to the contrasting terminological preferences of different clients. It also ensures consistency in terminology when many translators work on the same project.

Besides the aforementioned error typology, other types of translation errors can occur during the review of a translation, but those fall outside the scope of a TM's translator-assisting functionality. Such checks include:

- the appropriate preservation of meaning,
- the correct use of terms in a particular context,
- adherence to client-specific requirements regarding style and register
- other quality metrics as described for example in the J2450 Translation Quality Metric [Godden, 2000]

which must be left to the translator and other language professionals to detect and correct, although a TM system could certainly provide tools for the user to correct and record any such errors.

Since quality assurance is such an important task in any translation process, usually involving more than one person (depending on the size of project and its review cycles), a separate quality control module is essential for consolidating all quality-related functions and for independently assisting reviewers, proofreaders and QA staff in a translation department or company. In our survey, reviewers and proofreaders expressed the complaint that the translation review process does not fit well in the workflow of TM-assisted translation. They suggested that a QA toolbar would be a very practical feature in the editor interface. They also seemed to favour a graphical representation of errors and the activation and deactivation of the different types of warning and suggested correction. Most prefer the system to highlight any errors it finds from within the editor interface, showing everything in context, while others find more convenient the ability of the system to produce an external report showing a list of errors, allowing the user to navigate between the list and the files in which they are contained. The possibility of adding comments anywhere in a reviewed document is also favoured by translators and reviewers alike.

5.1.8 Collaboration

More than in any other sector, projects in the language services industry regularly require distributed operations and outsourcing. Some languages are translated internally, while others are passed on to service providers, who in turn assign them to freelance translators. From the point of view of translation managers, coordinating large multilingual, time-critical projects can often be a challenging prospect, with the need for intermediate synchronising of translation memories being just one aspect to be taken into account. Translators, on the other hand, increasingly start to think of themselves as participants in a multi-agent translation process, as revealed in our survey from their requests for features that enable collaboration.

Collaboration is not only about sharing resources. It is also about controlling and managing the fragmented translation workflow in a simplified and optimised manner, as well as enabling more effective communication between all the stakeholders in a translation project. TM-based translation support, in that respect, is expected to facilitate teamwork and help overcome some of the collaboration-related issues, resulting in reduced workload and cost.

The interoperability of TM systems is a major concern for all TM users, as revealed by our survey. Translation company owners and translation managers want to be able to share their translation assets freely and without problems, without being tied to the proprietary formats of their vendors and their TM system. Project managers need to be able to distribute their TM resources to freelancers without worrying whether the freelancer's preferred tool is compatible with the format of their resources. Translators, editors and

terminologists want, finally, to be able to exchange resources with their colleagues easily and, most importantly, to work on any translation project they wish to take on, irrespectively of the format of the resources supplied with it. As one freelance translator could not help but protest in our survey: *"I find it highly frustrating that I need to obtain various tm tools for various clients..."*

Unfortunately, most TM systems today suffer from a certain degree of autism, as they make use of proprietary file or database formats which cannot be perfectly interchanged. This is a fact that our survey's respondents find not only unhelpful for them personally but also harmful to the interests of the industry. All types of TM users, without exception, clearly need better ways of exchanging resources and seem to rest their hopes on the use of open-source formats for storing and exchanging resources. As for resources that have already been locked into proprietary file or database formats, they expect to see an improvement and the subsequent support by TM systems of the existing XML-based industry standards for the storing and exchange of resources, such as TBX and TMX for terminology databases and translation memories, SRX for segmentation rules, and XLIFF for translation files.

Communication between all team members in a translation project is also a need revealed by the respondents' requests. A dedicated built-in module would be helpful in improving communication and making its records available for future reference. In fact, it is often the case that other co-workers already have the information that new translators are asking for, and enabling them to access the company's knowledge base can both save work and ensure a prompt response.

In terms of translation workflow management, the suggestions, hence the needs, of the survey's respondents differed between those of translation agencies, and those of virtual translator networks and individual freelancers.

Translation companies and translation departments within organisations can find a range of connectivity options suitable for them, such as access to the server from the internal network (LAN), access over the Internet, server-to-server connections, and browser-only access to a remote database of resources, depending on the needs of their translation teams and their clients. Some desirable web-based collaboration methods that were mentioned were:

- centralised TM and working, whereby everything is controlled by the system and team members log in, allowing several language professionals to search, share, view and update the same repository simultaneously in real-time, or
- centralised TM with project-level downloads whereby all data can be downloaded with each new project.

A web-based platform is believed to facilitate the interaction between the translation company, the client, the translator, the proofreader and the in-country reviewer and allow

translators anywhere in the world to connect to and simultaneously access TMs residing on the server from their client version of the TM system. The greatest advantage of such a method is that freelancers do not have to own a TM system because they work directly in their client's web-based system. Sending projects can be straightforward and translations can be synchronised between people who work on the same project at the same time. An essential parameter of the configuration of such a system would be allowing the translators or reviewers to work either completely online or to combine offline and online working within the same software setup, depending on their internet speed. In addition, because several translators may simultaneously populate the same TM with their translations, the workflow engine must provide some type of protection mechanism for the quality of TM data, like for example an option which ensures that translations are first only stored in a virtual project TM before being transferred to the central TM after a quality check by an editor.

The development of TM servers is increasingly gaining popularity among corporate users of TM systems, as is also shown in the LISA 2004 survey [Lommel, 2004: 17-9], where respondents considered TM servers as a very important development in TM technology.

Our survey also encountered a need that has taken shape very recently. A few of our respondents reported performing translations voluntarily, as part of a community of volunteer linguists interested in some topic or product who want to disseminate information about the topic or product in different languages. A notable example of this trend is the open source communities who offer their language skills to localise open source software, to be used freely around the world. Incidentally, a related concept, that of "crowdsourcing" [Howe, 2008], has emerged and has been utilised by corporations who develop products and invite the users to localise them voluntarily, instead of assigning the localisation to specialised vendors (a famous example of crowdsourcing being the launch of the popular social networking site Facebook in several languages).

The needs of those translators concentrate on the ability to work and communicate effectively in a web-based collaborative translation environment. New technologies are needed to optimise the translation process, speed it up and make it more efficient. This relatively new phenomenon of collaborative translation has now been increasingly noticed by TM researchers and has inspired some interesting system designs [see Bey et al., 2006 and Murata et al., 2003].

Individual freelancers are also in need of groupware features that enable them to work with colleagues on a project. A few them, actually, seem to favour the idea of accessing server-based TMs via the internet (as in the example of Wordfast's Very Large Translation Memory, or Lingotek's 'public index'), in order to collaborate and share resources with colleagues in real time.

The general consensus among all the above users of TM systems seems to be that the Internet will become an increasingly important aspect of translation workflow, which is reflected by an increasing demand for Web-based solutions. The frequency with which networking features are requested can also reveal a tendency towards a significant shift from the use of standalone TMs to internet-enabled TMs, as workflows can be simplified and data synchronisation problems are avoided.

Chapter 6: What are the actual needs of TM users? – Part II: Non-functional aspects

In the previous chapter, we discussed the needs of translation professionals, as they derived from the responses to our survey, concerning the functionality, i.e. the operation, of a TM system. This chapter concentrates on the non-functional aspects of the system and needs in that respect. The understanding of non-functional aspects by software developers is just as important as correct description of functionality and plays a critical role in the design of quality software. Unlike the functional aspects, which describe specific concrete functionality, the non-functional ones refer to the general properties that the system or parts of it should exhibit (take, for example, the requirements *"the analysis of source files must be fast"* or *"I'd prefer a more user-friendly and intuitive interface"*) and are harder to quantify. The responses of our TM users revealed needs relating to several non-functional aspects, including usability, efficiency, portability, extensibility and reliability, as we shall see below.

6.1 Usability

The concept of usability in the context of TM systems denotes the ease with which people can employ the system as a tool to carry out any translation-related task. Furthermore, it

refers to the clarity and elegance with which the interaction between the system and its user is designed. Since the focus of this research is on the user and the optimisation of the use of TM systems as tools, needs in this area are particularly important. For the sake of clarity, they are broken down into more specific usability aspects of a software application: its user interface, its customisability, its learnability, and its supportability. These are drawn from the software characteristics taxonomy that was used in classifying the discovered needs (see 4.2.3 Classification scheme).

6.1.1 User interface

TM systems are interface-intensive applications in the sense that they need to be able to offer a rich visual environment where the available functionality and the various options are clearly displayed, and in which users feel comfortable working. Whether TM users employ a graphical user interface or a web-based interface, they attach great value to the experience that the interface of the system creates for them.

The translators who responded to our survey favoured the idea of a customisable layout in the translation environment, with at least the following visual components: menu bar, main tool bar, tabbed view for project, resources and search areas, source area with its tool bar and target area with its tool bar. The source and target area should be able to synchronise and allow the user to scroll them down simultaneously. All areas should be removable, resizable and displaceable. Search areas should be able to close automatically after use. In addition to these, one should be able to change the contents of the areas depending on what one needs to do while working. The translator toolbar must be powerful and receptive to the translator's needs by allowing her to select and centralise all the activities and functionality that she commonly uses.

In general, all TM users appeared to favour an environment where they could find all the information they needed displayed immediately as they worked without having to navigate around the product. They wished to be able to customise the interface with the windows that they needed, thereby clearing cluttered user interfaces, and to change the default display options and system settings quickly and easily.

The following needs emerged from the comments of respondents on user interface problems and suggestions for improvement:

1. Clearer user interaction:

Writing help messages tightly and making them responsive to the problem is considered crucial for comprehension and efficiency. The same applies to well-explained setting options. If something goes wrong or if something is outside the capability of the software, error messages should be clear, simple, comprehensible and explanatory of the situation, so that the user understands what is going on and what caused the problem. This is crucial for

troubleshooting the problem. Quite a few times, users reported: *"it did this... with no apparent reason"*. That is how users get frustrated and give up on a piece of software. Along the same lines, unclear or ambiguous messages about interactive tasks can cause confusion (a typical example of a poor attempt by a popular TM system to communicate with its user is the one offered by one of our respondents: *"I am flummoxed time and again by the following question stated in the negative form: Do you not wish to save the TM? Yes/ No"*).

2. More ergonomic user interaction:

Many of our respondents complained about not being able to use keyboard shortcuts to do things quickly (like entering terms to a glossary, or inserting a suggested match into the translation) and to avoid the constant use of the mouse which can harm one's wrist joints:

> *"[Name of TM system] is what I call a 'clicking nightmare'. One has to click all the time when working in [name of TM system], which is a real danger for our hands. I'm starting to suffer from RSI and I am convinced [name of TM system] has speeded up the process."*

Some consider the lack of support for shortcuts *"absurd for a text-processing application"*, and would be relieved to see interface dialogs designed for keyboard-only navigation and operation. The possibility of using keystroke shortcuts is also thought to increase productivity during editing and operating the program itself.

3. Visible navigation:

TM users seem to appreciate a navigation that is clear and natural, with the help of which they quickly see their range of options, grasp how to achieve their goals, and do their work. The navigation should present the illusion that users are always in the same place, with the work brought to them. This not only eliminates the need for maps and other navigational aids, it offers users a greater sense of mastery and autonomy.

4. Reversible actions & allowing 'undo':

People explore in ways beyond navigation. Sometimes they want to find out what would happen if they carried out some potentially dangerous action. Sometimes they do not want to find out, but they do it anyway by accident. By making actions reversible, users can both explore and make mistakes without worrying about damaging previous configurations. The possibility of undoing any action is also very important for correcting mistakes. The unavoidable result of not supporting 'undo' is that TM systems must then support a number of dialogs that say the equivalent of: "Are you really, really sure?", which is considered annoying and slows people down.

5. Use of colour:

Using colour in a TM interface is believed to enhance the aesthetic appeal of the environment and to make it resemble reality more closely. Indeed, a TM system that uses

red for terminology errors or highlights matches in yellow, for example, seems to imitate the behaviour of the translator as if she was translating on paper.

6.1.2 Customisability

When a TM system grants customisation facilities to its users, they can adjust the functionality of the system to suit their preferences and their work style. The need for customisation is apparent from many of the survey's responses (e.g. *"[I'd like to see a] Docking user interface that allows the users to configure their set up the way they like. Configurable fonts to make use easier on the eyes."*) but is also instantiated by the fact that quite a few users with programming skills have developed their own TM system, after failing to adopt a ready-made system [Lagoudaki, 2006].

Our surveyed TM users seemed to appreciate the ability to customise the functionality of all the tools offered by their system, such as the way to proceed through the text while translating, the terminology extraction, the text alignment and so on. In terms of the user interface, they would like to be able to customise the layout (for example, options for windows' positioning should include: horizontal/vertical, reverse, merged for image or paper translation, detach/attach, panoramic for large screens, panoramic for laptops), the colours and the fonts, as well as the keybord shortcuts. Furthermore, some users thought it would be helpful if a TM system allowed user options to be exported or imported, so that they can be used across different systems.

6.1.3 Learnability

The need for easy to use and easy to learn software may seem obvious to everyone, including TM designers, and one can assume that it has been at the top of their priorities. However, current TM systems seem to be suffering from a complexity that results in a steep learning curve, as shown by the overwhelming number of complaints registered by our survey's respondents:

> *"before I really understand what means a TM and how usefull it can be, it was really annoying to try to perceive how this whole idea (segmentation, storage of translation solutions, etc) can be usefull. Now I'm aware that the functionnality of TM's are not self-evident for neophytes ...they are not naturally attractive. And this is annoying: when something is good and we just don't catch how it can be usefull."*
>
> *"there are many functions which I can't remember how to operate. so for instance I never add words to the glossary now as I'd have to search the instructions for how to do it..."*
>
> *"Terrified of the technical problems associated with using the user-hostile software, incomprehensible manuals."*

> *"I used [name of TM system] almost exclusively for years and it took many months, with almost no support, to learn the rudimentaries. After 3-4 years I was using it to 10% of its potential."*
>
> *"It takes too long for you to get to learn how to use it and you never have time available to stop your work to learn, then you postpone and postpone it and end up never starting."*

Unsurprisingly, the difficulty of learning how to use TM systems was also mentioned as one of the most common reasons for not using a TM system among non-users. At the same time, learnability was mentioned by some as the plus factor about their TM systems when they were asked what they liked most about their TM tool. Evidently, the learnability of TM systems is of great importance to translation professionals, and issues with learnability seem therefore among the most serious problems that TM developers must address in the future. What users would like to see is software that is *"userfriendly, so you don't have to attend seminars to use them"*. In particular, users believe that extra attention should be paid to the simplicity of the user interface which should permit the user to start using the software right away without extensive training, but rather with the help of self-explanatory and comprehensible tutorials or screencasts. Users also expect TM developers to work on ways that enhance the learnability of their systems, such as offering better documentation and help, localising the software and its documentation into more languages, and avoiding the use of technical jargon for naming software features, tasks and actions.

6.1.4 Supportability

Supportability refers to the ability of the system to help users overcome problems when they occur. The inevitability of software problems is well understood among TM users: *"A minor bug is introduced with a new feature. Of course, this goes with the territory if one wants a very user-driven up-to-date tool"*. However, this was not the focus of our survey respondents' criticism. Their discontent seemed to start with the appearance of unhelpful, sometimes 'strange', error messages which gave no clue as to what was the real problem. As one user described:

> *"The single most annoying thing about [name of TM system] is that things occasionally go wrong, [name of TM system] grinds to a halt, and it can be very hard to understand what the problem is and how to fix it. This has happened to me dozens of times. So far, I have always been able to recover, and what's more without losing too much work, but it is a very hit and miss affair. The underlying problem to my mind is that the [name of TM system] documentation is not very good or comprehensive. That also applies to their support web site."*

More users seemed to agree that the manuals offered along with existing TM software packages are *"far too technical for a freelance translator who wants to keep it simple"* and that sometimes the "*help system is written for experts rather than 'mortal translators'*".

In the event of errors, lack of support from the TM vendor has been reported by many of our respondents, unless one buys a separate support contract. Some users also mentioned that their vendors' websites do not offer a substantial knowledge base where they could look for answers to their problem or even to simple practical questions they might have. The most serious shortcoming of the support for some of the available TM systems was believed to be the TM developers' failure to take into consideration frequent problems with the software, resulting in a distressing situation for its users where *"bugs remain unfixed year after year, version after version"*.

Evidently, TM users need to be able to find solutions to their software problems quickly and with confidence. Such confidence can be inspired firstly by good documentation and secondly by efficient customer communication and support responsiveness. Representative of this need is the response below by one of our survey's TM users:

> *"Better documentation, advice and support for handling problems and errors. [name of TM system] is a very complex product with a huge range of options, and it is often very hard to understand the benefits of different options. This is not helped by the [name of TM system] documentation, which I find often rather impenetrable. I speak as someone who has written a number of technical user guides in the past, and I would do it quite differently from the way [name of TM software company] do it -- for a start I would include far more example TMs to illustrate the various points. I would also include far more advice about how to recover from problems."*

In addition, there appeared a compelling need to have software documentation and help localised into as many languages as possible in order to decrease problems that derive from miscommunication. A possible solution to the serious challenge of localising the software's user interface and its documentation undoubtedly faced by TM developers could be crowdsourcing the translations to TM users/volunteers, by offering a Wiki-type method of compiling software-support knowledge bases.

Another proposal by a respondent that would probably enhance support responsiveness was the possibility of sending a message to the developer from within the TM system for either a persisting problem or a new feature request.

6.2 Efficiency

Efficiency in TM systems refers to how well they perform their funtions and at what cost in terms of computer resources. Our survey showed that each group of TM users had different priorities in terms of efficiency. For instance, terminologists were more interested in the efficiency of functions related to terminology collection or extraction. However, the majority of our respondents, who were translators, attached great importance to the efficiency with which the system leveraged existing linguistic assets. In fact, the majority of TM users seemed to be interested more in the performance of the tool in terms of matching,

than in the variety of features that a TM software package may offer. Efficient match recall and precision appeared to be their primary concern.

Efficiency in speed was also a concept that appeared repeatedly in our survey's responses. Some users pointed out that their system's speed and responsiveness appeared to "*slow down when dealing with certain files containing numerous or large graphics or have special formatting (text boxes, frames)*", or when the TM repository reached a certain size. Others seemed to be particularly irritated by the slow speed of crucial functions like match recall or spell check, which can slow down the whole translation process and drive users to think twice about applying TMs to their work, especially when it is urgent. A few TM users were also dissatisfied with the fact that their TM system used too much of their computer's RAM with the immediate effect of slowing down all processes. This did not seem to be of concern to those using web-based TM systems.

6.3 Extensibility

The ability of a TM system to be extensible, namely to allow its users to add functionalities to it or to modify the existing functionalities, is a need that appeared among translation company owners, as well as freelance users of TM systems with advanced IT skills. This segment of the TM market considers the benefits of real customisation as offered by an extensible TM application very important, as they believe they can achieve better software quality and greater reliability, have more flexibility, and reduce costs by not having to buy auxiliary applications. A TM system that supports plug-ins has many advantages for developers too. These include enabling third-party developers to create capabilities to extend an application, supporting features as yet unforeseen, and reducing the size of an application.

Extensibility means that the system is designed to include hooks and mechanisms for expanding/enhancing the system with new capabilities without having to make major changes to the system infrastructure. For example, a software system may have a public Application Programming Interface (API) that allows its behaviour to be extended or modified by people who do not have access to the original source code. This seems to be an ideal scenario for one of our respondents:

> *"One full-fledged app please, but make it modular, so you can keep tasks apart when necessary, and with a *well documented and powerful* API, so that add-ons like e.g. language-specific assembly algorithms can be made and plugged in."*

The most popular plug-ins for a TM system, according to the survey, were applications that translators believe can enhance their productivity, such as OCR, machine translation and speech recognition tools.

The most powerful way of rendering a TM application truly extensible is allowing access to its source code, so that its users can improve or fix the code as the need arises. Given that the development of open source TM applications seems to be on the increase [Lagoudaki, 2006], the demand for free applications is likely to become greater, particularly amongst those who want to adjust their TM system every time their needs change and those who resent TM vendor lock-in.

6.4 Portability

A frequent criticism by TM users, as appears from our survey, is that their TM system is not portable from one platform to another. This is particularly frustrating for people who are used to and enjoy working on a particular platform, but since they cannot use their preferred desktop TM application on it, they are forced to buy the platform on which their TM system has to be installed. Most frequently, this change affects the learning curve of the TM user and incurs additional costs. The underlying need of many in this situation is desktop TM systems written in a multiplatform language, so that they can be used seamlessly across all available platforms. Web-based TM applications, which inherently overcome this problem of platform dependence, should also be developed further.

With the rise of small-screen devices (like PDAs and smartphones), a need seems to be emerging for mobile versions of TM systems, so that translators can take texts with them and edit them anywhere. Although the user who expressed such a request believed that he/she *"would be a minority on that one"*, current trends of mobile working among translation professionals indicate that more requests of this kind are likely to appear in the near future.

6.5 Reliability

The reliability of a TM system is apparent from its ability to perform and maintain its functions in routine circumstances, as well as under hostile or unexpected conditions. Defects are inherent in every application, as we have pointed out earlier, and unexpected situations are bound to arise. According to Beck and Andres [2004:97] defects can destroy the trust of users. The goal of the development team must then be to reduce the occurrence of defects to a level where trust can reasonably grow, by investing in defect reduction strategies such as extensive testing and continuous user feedback.

Our respondents reported frequent occasions where their system crashed unexpectedly resulting in the loss of some or all the work they had completed up to that point. A lack of user warnings, error messages and tools seemed to compound the problem.

Most TM systems nowadays tend to use an Autosave function, meaning that one's work is saved in the translation memory and can easily be retrieved in the event of a crash. But that could still see someone losing several minutes' work in the event of a crash.

What users apparently need is a system that saves everything carefully and continuously to its translation memory, including saving each translation unit in a project as soon as the user proceeds to the next one. The system must ensure that users never lose their work as a result of error on their part, the vagaries of internet transmission, or any other reason other than the completely unavoidable, such as sudden loss of power to the client computer.

6.6 Addressing conflicting needs

TM users are a diverse population, in the sense that their needs can vary to a great extent. Needs diversity involves accommodating users with different roles, experience, computer skills, language combinations and specialisations, and it is one of the greatest challenges faced by TM developers. While a number of the existing commercial off-the-shelf TM systems seem to have avoided the pitfall of trying to accommodate every single requirement expressed by their users, others have not succeeded in this. Hence, we see on the market systems with an abundance of features that are useless for certain categories of users, or systems that in their effort to offer solutions that will benefit one user have placed constraints on another.

The risk of producing a false-feature rich system is real for every TM system that is aimed at a wide range of user types (e.g. translators, terminologists, project managers). Such a system is most likely to suffer in terms of usability, as its features are not 100% deployable (some may be redundant and meaningless for some users) and the added complexity is likely to increase the learning curve of TM trainees. But how can TM developers satisfy such diverse users to an equal degree?

Temponi et al. [1999:346] observed that requirements (the same being true for user needs, as revealed by our study) can be *mutually exclusive*, *irrelevant*, *cooperative* or *conflicting*. Two needs are called mutually exclusive if they cannot be satisfied (partially or completely) at the same time. They can be irrelevant to each other, if satisfaction of one need does not have any impact on the satisfaction of the other need. Cooperative needs are considered two needs if an increase in the degree to which one need is satisfied also increases the degree to which the other is met. On the other hand, when two needs are seen as conflicting with each other, it means that an increase in the degree one need is satisfied often decreases the degree to which the other is met. An example of conflicting needs is shown below:

> *"Feature-rich does not mean overbearing. I would prefer a feature-rich, single small-scale tool that is highly-adaptable to individual translators' needs and that places preference on new features that enable a translator to get started*

> *quickly on any format (OCR, PDF handling), to access additional reference materials more directly (CD dictionaries, major online dictionaries, bilingual parallel texts), and to work from within any of the major word processing applications and OSs. That being said, I don't believe that any one tool provider can develop a single end-all be-all tool."*

The above user shows a clear preference for a variety of features that he/she thinks are important to his/her work: OCR, access to additional reference material and so on. But although for some users these features may be vital, for others they are too 'fancy' and unnecessary:

> *"The above 3 suggestions are very nice, but possibly a bit fancy. Can a TM tool do these any more easily than I can, and more importantly, should it? I'd prefer TM tools to concentrate on getting the basics right (why *can't* most others, apart from [name of TM system], Assemble, and what good is a tool that doesn't, for example?) before they start adding loads of potentially useful but inessential additional features (more development time, more cost to the user, more things to update)."*

The conflicting needs in this example are not the needs for particular features as requested by each respondent. Those can be cooperative or irrelevant. The real conflict lies in the need for enhanced and rich functionality on one hand and the need for simplicity and 'pared-down' functions on the other. In other words, when the developer attempts to please the first user by adding additional functionality, the satisfaction of the second user's need for simplicity of is likely to decrease.

Based on our investigation of user needs, conflicting needs usually result from differences in the use of context by different user groups (e.g. translators, terminologists, project managers), from a variation in levels of experience both in translation and in using the system, and finally from differences in the users' level of computer usage competence.

So how can TM developers reconcile divergent needs? An idea suggested by one of our respondents is the distinguishing and grouping of functionalities according to translation roles under separate modules and the availability of a dashboard that can give easy access to all software modules. This will allow any type of user to start her project from the module that is most relevant to her task without being presented with features that she is not likely to need. Another suggestion, more suitable for corporate systems or collaborative environments, involves the presence of customisable user levels. In each user level, certain features are either enabled or disabled, based on what would be more useful for each user who logs into the system, with even the potential for tailoring the interface to each user's preferences.

The needs of TM users do not only occasionally conflict with each other. They can also conflict with the needs of TM developers, which are sometimes considered of higher priority, and as a result the conflicting need of a user is automatically overlooked. A striking example of such a conflict of needs is found in the development of linguistic data that would

help increase the efficiency of functions such as text alignment, match assembly and terminology extraction.

Based on our investigation, translation professionals do not need tools that function for *all languages* but tools that work well for *their* working languages. In an effort to address such needs, TM developers would have to develop (with substantial cost and human resources involved) linguistic data in every language combination, which is economically forbidding as a solution. If, alternatively, they developed linguistic data selectively, namely built TM systems that work only for a few language combinations, as happens with some of the existing commercial TM systems, in an effort to address the needs of users in major segments of the markets they serve, they would satisfy some users. However, at the same time this approach would leave many feeling discriminated against and seriously dissatisfied. Moreover, building language-specific TM systems does not only show a disregard for valid user needs but also creates a backlash for TM developers. In fact, by adding linguistic knowledge to a TM system, the product automatically loses its language-independence and its market share shrinks dramatically, especially since there are many translation companies that deal with multiple languages.

A possible solution proposed by some of our respondents would be the availability of global TM systems and the development of language-specific add-ons (linguistic data and tools), that can be sold separately, based on a careful cost-benefit analysis that takes into account the high demand for particular language combinations. Obviously, such a solution would leave minority languages in a disadvantaged position, as the low demand for rare language combinations would not give sufficient incentive to develop costly language resources. In this case, TM developers offering MT capabilities should perhaps direct their research efforts to ways of acquiring lexical resources quickly, easily and cheaply from sources that contain readily available multilingual corpora. An enormous source for such material could certainly be the Web.

Add-ons could also be developed for additional functionality if requested by experienced users, whereas novice users and those who preferred using only the essential functions could still enjoy the basic version of the system. The solution of add-ons seems to be useful in a market that is characterised by heterogeneous needs, as users are able to customise their application before and after they install it by mixing and matching the functionality that they want.

Conflicts in user needs are bound to appear in any requirements engineering process, but what is important is how TM developers deal with them. Some of the needs may be truly unnecessary; others may turn out to be temporary interests. The developer must resolve the conflicts in collaboration with the users, making sure that various categories of user are involved. Together they should delineate which needs are of high priority, which are actual,

which are potential, which are distantly related, which may be dropped and which should be added to the list of forecasted needs.

In the light of such a diversity of needs, it seems a hopeless task to look for an answer to a general and inaccurate question such as "What is the ideal TM system?". It is clearly impossible to accommodate every individual's preferences and desires within a single system, in the same way that it is impossible to identify a prototypical TM user. What is important, however, in such an atomised translation society, is free market and the availability and constant development of various different TM systems, so that each individual user has a better chance of finding the system that matches her needs most closely.

Chapter 7: Knowledge resources and innovation in TM systems

In the previous two chapters we discussed how the functionality of TM systems can be expanded and how the overall quality of TM software solutions can be improved, based on the needs distilled from the responses to our survey. Yet, needs can lead to something more than micro-optimisation. Some of them may even hold the key to innovation for translator-support technologies.

It is widely acknowledged in the realm of product design that opportunities for novel systems or features can be either *created* through technological innovation, or *spotted* through needs. Most TM developers follow a holistic approach to their innovation sourcing strategy which considers customer needs and technology trends, and aims at matching customer insights to systems' design. They are particularly interested in identifying 'lead users' who according to Von Hippel [1998] can be defined as those who are early adopters and passionate users of company products, and work with the company to drive innovation. Insight from 'lead users' is thought to be an excellent indication of what the customer base will need in the near future. Usually, this group of users consists of corporate users of TM systems, such as translation managers and translation company owners, as they are arguably the first to adopt TM solutions; therefore their experience in using a particular TM system outgrows that of freelance translators, for example, who are considered a late and slow adopter of translation technology. Corporate users of TM systems are also preferred as

'lead users' because they can test their systems in complex and large-scale translation operations and, subsequently, offer rich feedback in terms of desirable functionality.

István Lengyel [2006] in his article in the leading industry magazine *MultiLingual* looks at innovation in TM systems from the same point of view as that of corporate users. He believes that the source of innovation is to be found in technologies that offer a better control of the workflow in translation projects as the latter become increasingly fragmented. The findings of our survey have allowed us to offer an alternative source of innovation, looked at from the perspective of freelance translators who were the majority of the survey's respondents.

Innovation sourcing, in the context of this research, has also been based on the insights of 'lead users'. However, in the case of freelance translators, 'lead users' or 'power users' were considered those who appeared to be the most intensive users of a TM system, and the people who used the system to the fullest. Certainly, not all needs expressed by 'power users' qualified as containing seeds of innovation. But some needs appeared to be 'exciting'.

The term 'exciting needs' was first introduced by Kano [1984] who argued that in order to increase customer satisfaction, product designers must distinguish and incrementally meet three types of need: *basic* (which concern *expected* features), *performance* (which reveal *desired* functionality) and *exciting* (corresponding to features that are *well beyond the imagination or expectations of the customer and cause excitement when implemented*). Customer satisfaction with the application can be dramatically improved through the implementation of a few 'exciting' needs. Moreover, 'exciting' needs, if fulfilled, provide important competitive advantages for the developers and inspire loyalty in customers [Griffin and Hauser, 1993:8]. However, failure to implement these needs does not adversely affect overall satisfaction [Krogstie, 1999].

According to Kano, 'exciting' needs are never fully articulated as they have not been crystallised yet in the minds of the customers (TM users in our case). However, we can anticipate them by studying the situation of the user and by identifying what are the problems and opportunities arising for this situation. The resulting understanding of the goals of the user allows us to explore what new benefits the user would appreciate. From the new benefits, we can then look for product attributes that would contribute to those benefits. Such features should be exciting to the user.

One 'exciting' need was identified within the following TM user's response:

> *"I would very much like to be able to centralize the content of all the dictionaries I have purchased and of glossaries available free on internet, in order to query against them as I translate and have the possibility to insert translation of terms in my TM environment."*

Among his/her other needs that are clearer to identify (such as the need to consolidate the language resources in one location, the need for simultaneous search and the need for

inserting text from external documents into his/her translation editor), we can anticipate a need for dynamic interaction between a TM system and the internet. This user, according to supplementary data on his/her profile, has a tendency to use online resources when facing translation problems. Specifically, he/she reported that his/her translation research methods included first searching on the internet in online dictionaries/glossaries, then in dictionaries/glossaries on CD-ROM, then in hardcopy dictionaries/glossaries, then again on the internet, via a search engine (e.g. Google) and in monolingual texts on the subject, and if no solution can be found in the above sources, he/she then finally resorts to asking a colleague, or a subject expert.

From the answers to the same question regarding popular translation research methods, a predilection for online resources is apparent amongst all the translators in our survey. The table below shows the most popular research methods and how they were ranked by the survey's respondents:

Research method	**Ranked**
look in dictionaries/glossaries on CD-ROMs	1st
look on the internet: in a search engine (e.g. Google)	2nd
look on the internet: online dictionaries/glossaries	3rd
look in hardcopy dictionaries/glossaries	4th
look on the internet: bilingual (parallel or aligned) texts on the subject	5th
look on the internet: monolingual texts on the subject	6th
look in old translations/glossaries that you have	7th
ask a colleague	8th
look on the internet: ask questions in forums/communicate with other translators	9th
ask a subject expert	10th
look in printed texts on the subject	last

Table 11: Popular translation research methods by order of preference

This evidence suggests that translation professionals make very frequent use of resources, especially those found on CD-ROM and on the internet, that lie outside their TM-enabled workflow. Bringing these resources closer to the translator is anticipated to have substantial benefits. It is believed that such an 'exciting' need that is currently latent has the potential to bring innovation to TM systems.

Another response that echoes the same 'exciting' need was the one given by another 'power' user:

> *"I've just finished working on a geological report (nearly 20,000 words rendered from Brazilian Portuguese into English) for which I had to use a huge amount of my time on the internet (Google) to make sure the appropriate*

> *wording would be used (I am an aeronautical engineer, not a geologist!). If only I could have a user-friendly tool to help me in such situations..."*

It is worth noticing, at this point, how this particular need reflects what Austermühl [2001:38] has described as a 'paradigm shift' in the research methods of modern translators (namely, a shift from hardcopy reference material to digital and online resources), which coincides with what O'Hagan and Ashworth [2002:142-156] have called the 'new paradigm of teletranslation'. Research by Fulford and Zafra [2004:58] on the uptake of online tools and resources by translators also reveals that modern translators are making extensive use of internet-based facilities to assist their research, especially the younger less experienced translators who have received some type of formal education on translation. The reason for the increasing uptake of internet-based facilities is well captured in Asensio's observation [2003:111]:

> Professional translation, as a profitable activity, imposes restrictions on the effort that can be invested in finding information. Researching *ad libitum* may be characteristic of translating as an ideal mental process, as a scholarly activity or as a part of the training stage. In professional activity, however, we simply do not have unlimited time or resources. Obviously, the internet is now our most powerful tool, giving quick access to countless sources of information.

The importance of digital resources for translators can signify an opportunity for TM system developers who wish to extend their systems' functionality. Up until now, based on the available commercial systems, most R&D efforts in TM technology seem to have been focusing in the maximisation of the utility of the translator's legacy resources (mainly by improving the matching). However, considering the fact that a translation memory is easily saturated when it is small – preventing any sophisticated matching to occur at peak performance – we tend to believe that innovation in translation support is more likely to come from expanding the translator's resources, either by pointing translators to relevant resources and allowing access to them from within their TM system, or by integrating sources of information into the TM system. The development of tools for the acquisition, access, management, customisation and use of resources is anticipated to bring substantial benefits to the investigative nature of a translator's work as such tools will help translators open up to additional sources of knowledge.

7.1 Access to knowledge resources

The importance for translators to be able to access large lexical resources online is increasingly recognised not only by TM developers (see for example the Very Large Translation Memory project[43] by Wordfast, thanks to which users of the Wordfast TM system can have access to an enormous public translation memory), but also by other

[43] More information can be found on Wordfast's website at: http://www.wordfast.net/index.php?whichpage=jobs&lang=engb [last accessed on 20/09/2008]

parties interested in translation technology. Google's recent announcement of the future launch of Google Translation Centre,[44] which will allow freelance and professional translators to search a vast collection of parallel texts, is yet another indicator of future directions in translation technology.

Translators seem to be interested in two types of knowledge resource: lexical (which help them find translation equivalents in the target language) and factual or extra-linguistic (which help the pragmatical disambiguation of their source text and the validation of their translation) [Kim, 2006:285]. Lexical resources can be translation memories, bi-texts, bilingual websites, dictionaries and specialised glossaries, whereas factual resources can be monolingual texts or websites relevant to the source text, monolingual dictionaries and encyclopaedias.

The need for TM systems that allow access to all the resources that a translator usually consults from a single platform, regardless of where the information may physically reside, is soon to be even more apparent. Specifically, translators would benefit from a system that can send queries for a source term/phrase simultaneously not only to local TM resources (as all TM systems do now), such as the translation memory, the terminology databases, archives or glossaries, but also to:

- CD- or hard drive-based dictionaries and glossaries
- online search engines
- online resources

making it possible to link information stored in these distributed, heterogeneous environments.

The user should be able to preconfigure the offline dictionaries/glossaries and the online search engines that the TM system sends the queries to and he should be allowed to bookmark new search or metasearch engines. He should also be allowed to create his own search engine using only the sources he trusts, i.e. specify the websites that he considers most authoritative, from which the search engine should provide results.

New sources of information, such as electronic journals and e-books, represent a good opportunity for harvesting suitable knowledge. A link to the online open encyclopaedia Wikipedia would also be very useful for many translators for retrieving encyclopaedic information and also for identifying translation equivalents since many of its articles can be found in other languages too (even though these different language versions are not translations of each other). Interestingly, a study by Jim Giles [2005] has shown that Wikipedia's status as a resource for factual information is rising to the same level as Encyclopedia Britannica in terms of trustworthiness, which offers some reassurance to the quality-conscious translators.

---

[44] For further information, read the article by Brian McConnell at: http://gigaom.com/2008/08/04/google-translation-center-the-worlds-largest-translation-memory/ [last accessed on 20/09/2008].

The TM system could also allow users to bookmark any external websites which contain glossaries of interest and even to store these for offline search. Certainly, the system must also offer information on these sources to allow translators to assess their trustworthiness, as one of our respondents pointed out:

> *"A translator needs to mistrust any terminology from a remote or unknown source, so the source needs to be kept apparent when you start integrating remote terminology resources".*

In the same spirit of bringing translators closer to more information and expert assistance, a TM system could integrate fora, blogs, translators' portals and vortals. As a closer cooperation of translators through online communities is predicted for the future [Désilets et al., 2006; Bey, 2006] a system that favours interaction is likely to bring considerable advantages to translators, especially to those who feel that the translator's job is unpleasantly isolated.

7.2 Acquisition of knowledge resources

Not all translators can or want to work online, while some prefer to own their knowledge resources, feeling safer in knowing that the resources' quality has been validated by them personally. This realisation has sparked an interest in trading quality knowledge resources[45] among those who have generated massive translation memories over the years from large-scale translation operations and those who enter the translation arena with no substantial legacy memories or those who simply wish to expand their knowledge resources in an easy and quick way.

Such knowledge resources, as well as dictionaries, glossaries and idiom lexica (for assisting in the translation of general texts), could be sold by TM developers in the form of language packs, something that would certainly add value to their main TM product. TM developers could also collaborate with established dictionary publishers or producers of multilingual content (such as accredited newspapers or other media companies) to offer integrated translator-support solutions (TM systems with language-specific resources) from which translators could greatly benefit.

7.2.1 Bilingual corpora

In the array of knowledge resources that translators frequently consult, bilingual corpora feature in a prominent position. There can be two types of bilingual corpus: collections of a) *comparable* texts (texts dealing with the same subject in two different languages) and b)

[45] See for example the TM marketplace initiative, a virtual marketplace for selling and buying translation memories: http://www.tmmarketplace.com/ [last accessed on 21/09/2008].

aligned or *parallel* texts (original texts with their translation). Comparable texts can be very useful to translators for identifying indirect contextually appropriate equivalents of expressions that appear in the source text [Babych at al., 2007]. Because translators are not necessarily subject-field experts, they need to consult texts originally written by experts in order to make sure that their translation adheres to the terminology and style being used. This way, the quality of their translation output is usually improved and consistency in the use of domain-specific terminology is easier to maintain.

Parallel texts are also seen as a multi-purpose resource with enormous potential for assisting the translation activity. First of all, what seems to be the most obvious use of a parallel corpus is that, depending on the size and subject-domain of the corpus that translators have at their disposal, they can compare it with their translation project to find any existing translations of parts of their text. From a parallel text corpus they can also extract bilingual terminology with the help of term extraction tools, and build specialised glossaries [see Somers, 2001]. In fact, Bowker [2004:77] points out that when a translator has a bilingual parallel corpus it is like having a sort of "full-text glossary" which by its very nature contains countless "term records".

Additionally, parallel texts can be used to validate the translator's choices, by cross-checking them with word frequency data or by being able to see the chosen terms in a variety of contexts simultaneously [Austermühl, 2001: 124].

Finally, parallel texts can be used for different types of text analysis either for investigating translation strategies or for acquiring or cultivating linguistic or subject-specific knowledge (see studies by Laviosa, 2003:105-117; Zanettin, 2002:239-248 and Laviosa, 1998:1-6 for the pedagogical role of corpora). Evidently, such corpora give the translator the ideal opportunity to discover underlying patterns for the two working languages and to obtain an in-depth overview of all kind of equivalences (semantic, stylistic, structural, pragmatic) that exist between two parallel texts.

Despite the fact that parallel text corpora are considered such an important resource, only a few are available for public use (due to copyright restrictions) and those are limited in size and language coverage. Some of the biggest aligned corpora[46] (in electronic format) that have been built so far and are publicly distributed are: the Hansard corpus[47] (English-Canadian French/ 2.87 million parallel sentence pairs), the Europarl[48] (11 European languages/ 44 million words), the UN corpus[49] (English-French-Spanish/ 165 million

[46] An extensive list of parallel corpora can be found at David Lee's website: http://devoted.to/corpora/ [last accessed on 17/09/2008]
[47] Collection of official records of the proceedings of the Canadian Parliament from the mid-1970's through 1988, available at: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC95T20 [last accessed on 17/09/2008]
[48] Collection of proceedings of the European Parliament from April 1996 to October 2006, available at: http://www.statmt.org/europarl/ [last accessed on 17/09/2008]
[49] Collection of documents offered by the Office of Conference Services at the UN in New York; they are drawn from archives that span the period between 1988 and 1993; available at: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC94T4A [last accessed on 17/09/2008]

words) and the recent release of the European Commission's Directorate-General for Translation (DGT)'s multilingual *Acquis Communautaire*[50] (the entire body of European legislation in 22 languages).

What is promising is that the number of parallel texts is constantly increasing and large amounts of open content are being created (e.g. Wikipedia, news, blogs, wikis, software documentation[51]), but most efforts are still concentrated on majority-language pairs. A further disadvantage of the currently available parallel corpora is that most of them are unbalanced due to the fact that they represent mostly governmental or newswire-style texts. This means that even though they all might have the same style, they fall into a wide range of subject domains without this signifying that they are a representative sample of a particular domain. And of course, as most things that require considerable time and effort to create, parallel corpora usually do not come free of charge; their acquisition is frequently accompanied by fees and licensing restrictions.

7.2.1.1 Compiling parallel corpora

For translators, building a large parallel corpus is a major issue. Faced with the need to have a parallel corpus, they are presented with one, or at most two, options: either to access any of the previously mentioned ready-made corpora (if there is one available for their working language pair) or to build their own private collection. The private collection is the most commonly followed route, according to many translators. In fact, our survey revealed that only 12% of translators use ready-made corpora, whereas the vast majority are accustomed to compiling their own corpus as they translate or by aligning their legacy translations.

But despite the availability of numerous tools, such as concordancers and alignment tools designed *inter alia* to enable and speed up the corpus compilation process, the fact remains that corpus compilation is a very time-consuming and tedious task with rarely satisfying results.

A few commercial TM system developers have recently recognised the importance of being able to build a large parallel corpus fast and with little effort. For instance, MultiTrans (by MultiCorpora) and Lingotek (by Lingotek) have applied a full-text approach to TM database creation, which allows translators to build their database rapidly by importing and indexing the full text of their legacy bitexts, and by also skipping the sentence alignment process[52] that is initially required by all other, sentence-based, TM systems. Unlike most TM

[50] The DGT Multilingual Translation Memory of the Acquis Communautaire boasts of being the biggest parallel corpus in existence so far. Its most outstanding advantage - apart from being freely available - is the number of rare language pairs (e.g. Maltese-Estonian, Slovene-Finnish, etc.). It is available at: http://langtech.jrc.it/DGT-TM.html [last accessed on 17/09/2008].

[51] Particularly encouraging is the free culture movement promoted by Creative Commons, thanks to which an increasing number of old and recently authored texts are being released from their copyright and become available in the public domain. For more information see http://creativecommons.org/text and http://www.opencontentalliance.org/index.html [last accessed 20/09/2008].

[52] Both systems align the parallel texts on paragraph level instead.

databases, the corpus in the two TM systems does not depend on perfect alignment, so this eliminates the need for time-consuming up-front verification of the aligned SL and TL pairs before a translator can start using his/her corpus. But even though they offer a way of speeding up the creation and exploitation of a parallel corpus, they assume that a generic form of such a corpus already exists, i.e. that translators have at their disposal a large number of original texts with their translations, or that a generic corpus can be manually downloaded from the web. More often, the first assumption does not hold true, because very few translators, even those with many years of working experience, have at their disposal a massive organised private collection of parallel texts in electronic format. As for the second assumption, downloading parallel texts from the web is indeed a possibility, but bearing in mind the vast number of web locations that contain parallel texts, it is unreasonable to believe that manually locating and downloading pairs of parallel texts one-by-one could be an effective way of building a large parallel corpus.

When translators are creating a parallel corpus they aim to collect as much textual data as possible covering a wide range of domains with the minimum effort and time spent on the task. They seek parallel texts according to criteria such as the following:

- Texts in electronic formats
- Publicly available texts with 'loose' copyright and ownership restrictions
- Texts from authoritative sources
- Texts representative of a specific subject domain
- Texts with terminological relevance to their translation projects

Luckily, there is already a single source where vast amounts of such texts can be found: the web. It is unquestionably the largest collection of electronic texts and as such it presents an excellent opportunity for exploitation.

7.2.2 The web and its characteristics as a repository of texts

The web has some unique characteristics that distinguish it from all other compiled (printed or electronic) corpora. First and foremost, its size (indicatively, over one trillion public webpages have been indexed by Google[53]) surpasses the size of any large electronic corpus and is steadily growing day by day. Unlike structured corpora which remain static after they are built, the web, thanks to its dynamic nature, retains the advantage of always offering 'fresh' content, thus allowing translators to trace the ongoing evolution of a language according to its latest usage trends. Also, the fact that there is so much publicly available material that one can download for one's own private use without having to pay for it is an additional advantage worth considering.

[53] Data from the official Google blog: http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html [last accessed on 16/09/2008].

Finally, one can observe that the dominance of the English language on the web is starting to recede, as content becomes truly multilingual because more languages are being used. Even minority languages are increasingly represented on the web, which gives us reason to believe that we can find a considerable amount of parallel text for 'rare' language pairs.

Although everyone recognises the superiority of the web as an immense repository of texts, many translators, as it also appeared from our survey, question the quality of such content and are reluctant to use it as a corpus for their work. They argue that inappropriate web content not only has little utility as background information but can actually interfere with the translation process and lead to the production of translations of poor quality.

Admittedly, there is no guarantee that a corpus collected randomly from the web will be of 'good quality'. But then again what is meant by 'good quality'? Is good quality language that used by scholars, or educated people in general? Or the one that obeys the rules of grammar and has a 'correct' syntax? If that is what we have in mind in terms of quality, then the web is not a resource that can be used arbitrarily without pre-processing. We will have to restrict the search for bilingual content according to our 'quality' criteria. But at the same time, we need to be aware of the discrimination we commit against language, if we proceed in this way. The language used on the web has one attribute thanks to which it can be considered the finest sample of human language: it is free and highly expressive. The web as an open linguistic resource is able to expose plentiful uses of a language, not only the 'correct' formal types but also the informal ones that are typical of oral discourse. This enormous variety of language phenomena that can be revealed is what constitutes the quality of web content.

7.2.3 Automatic extraction of parallel texts from the web[54]

By now translators have realised the potential of the web as a source of parallel texts, but they lack access to tools capable of extracting these texts quickly and on a large scale. So far, no commercial TM system features a module for the automatic extraction of parallel corpora from the web. However, on the academic research front there is some considerable activity around the development of such tools, which has produced some very encouraging results. Some well-known web-mining tools are: STRAND [Resnik & Smith, 2003: 349-380; Resnik, 1998], BITS [Ma & Liberman, 1999] and PTMiner [Kraaij et al., 2003: 381-419; Chen & Nie, 2000].

STRAND is the first web-mining tool that was developed to automatically identify and extract parallel corpora from the web. At its current stage of development, two approaches

[54] This (7.2.3) and the next section (7.2.4) have been published in a modified form in the proceedings of the X International Symposium on Social Communication, Santiago de Cuba, 22-26 January 2007. Paper title: "Challenges and Possibilities for Extracting Parallel Corpora from the Web – The Translator's Dream Scenario" by Elina Lagoudaki.

to the system's architecture have competed for better results: the first approach is the structure-based, while the second is the content-based. According to the first approach, the tool originally worked on the assumption that translated pages on the web tend to exhibit parallel structure and they follow the same rules of content authoring; therefore the tool was designed to look for those similarities. By using the AltaVista search engine's advanced search, the tool searched for any pages that include anchors related to the two languages one is interested in, and in this way it retrieves a list of candidate webpage pairs. After this, in order to check that the retrieved bilingual pairs are parallel, the web crawler component of the tool performed a URL matching to identify any similarities in the pages' respective paths. It then looked at the structure characteristics (HTML mark-up) of webpages to identify any parallel structure, until it finally produced a list of filtered out parallel webpages. On the other hand, according to the content-based approach, it is recognised that not all web authors follow a standard mark-up convention (such as XML) and some webpages do not even have any structural mark-up. Therefore, a modified version of the STRAND tool makes use of a word-to-word translation lexicon in order to identify similarities in the content of two paired webpages. The results offered by the structure-based approach (following a ratings-based assessment) were 100% precision[55] and 68.6% recall[56] for three data sets created from an English-Chinese corpus, whereas the content-based approach showed (through human evaluation) 97.4% precision and 98% recall for an English-French corpus of 325 webpage pairs [Resnik & Smith, 2003: 349-366].

Bilingual Internet Text Search (BITS) is also a tool that searches the web for parallel texts, but it does so through a list of domains, generated by a candidate generation module, which have a high probability of containing parallel content. It operates by sampling pages from each domain and identifying their languages; if a domain is found to be multilingual, all pages on the site are crawled, while if it is monolingual then the system moves on to process the next website. BITS considers all possible combinations of web page pairs in the two languages and filters out incorrect pairs by using a large bilingual dictionary to compute a content-based similarity score and by comparing that score to a threshold. In addition to cross-lingual lexical matching, BITS filters out candidate pairs that fail to match in terms of file size, anchors (numbers, acronyms, and some named entities), or paragraph counts. Using an English-German bilingual lexicon of 117,793 entries, Ma and Liberman [1999] report 99.1% precision and 97.1% recall on a hand-picked set of 600 documents containing 240 translation pairs (as judged by humans). This technique yielded a 63MB English-German parallel corpus [Resnik & Smith, 2003: 359].

Parallel TextMiner (PTMiner) uses a more advanced technique to identify candidate parallel webpages and verify their parallelism. Initially, it sends requests, specifically formed to include anchor texts identifying the language, to a powerful internet search

[55] By precision is meant the proportion of bilingual webpage pairs retrieved.
[56] By recall is meant the proportion of page pairs correctly identified as parallel among all bilingual webpage pairs.

engine (e.g. Google or AltaVista) and the websites retrieved are considered to be the candidate bilingual websites. As a further step and based on the fact that a website is a host corresponding to a distinct DNS address, it employs a host crawler that crawls each candidate site for more URLs that could contain bilingual content. Once a large collection of URLs has been identified, each webpage in each pair undergoes several tests in order to find out if the paired webpages constitute parallel texts. To discover this, the tool examines the file names and the URL paths of the webpages for similarities in the prefixes and suffixes for both the source and target languages. After this stage, the webpage pairs are filtered further by criteria such as file length, HTML structure and language and character set. Kraaij et al. report approximately 99% precision (manually checked) for an English-French corpus, while the final corpus extracted was 174/198 MB [Kraaij et al., 2003: 391-2].

All the above tools, despite the different techniques they use, have been able to extract from the web an impressive amount of data, and therefore, they provide evidence for the fact that it is feasible to compile parallel corpora from the web automatically in a short time and with little human effort.

7.2.4 Integrating a corpus extraction facility into a TM system

During our survey, translation professionals were asked to assess the idea of a Translation Memory module that "locates bilingual parallel texts on your subject on the web (you can specify authoritative sources only), downloads them, aligns them and fills in your TM database for future use or reference". The reaction was overall positive, with the majority of respondents (75%) finding a tool that offers such functionality very or somewhat important to have, provided that they would be able to customise the content search to ensure quality.

The reason why a web-mining tool should be integrated into a TM system rather than being developed as a separate utility is because the benefits deriving from the use of such a tool will be maximised if combined with the advantages of a TM system. For instance, it is not enough if translators have a huge parallel corpus in their database; they also need a search and retrieval tool that will allow them to exploit this corpus. Also, given that TM systems are the tools most commonly used by translators compared to other available CAT tools, it is more practical for translators to have one single tool for doing their job rather than having various different tools that cannot interact. Furthermore, many of the modules of the TM (such as the database and/or the alignment module) can be shared with the web-mining tool, thus reducing the cost of developing such a tool. Finally, a tool that fills the TM repository automatically is something that has always been missing from TM systems and, at the same time, seems to provide a plausible response to the 'exciting' need of integration of resources.

In an effort to visualise such a tool as a module integrated into a TM environment, the model can be based on the example of WATCHER, an intelligent agent developed by Sgarbas et al. [2003:449-464] that is designed to extract multilingual corpora automatically from the internet. This particular tool exhibits some very interesting features that could be easily accommodated in a TM system, but it does not deal with parallel corpora identification.

The proposed agent would crawl the web looking for parallel content according to a combination of structure and content-based criteria used by all the aforementioned experimental tools. The process would include the following steps (some performed online and others offline):

Online actions

I. Location of pages that might be mutual translations: The web crawler would be able to identify language-specific webpages in every publicly available website by a) checking for language indicators in the URL paths or the file names and b) parsing the content of the webpages for any language anchors (e.g. 'English version', 'français', etc.). The results would form a preliminary list of paired webpages.

II. Cross-checking that a pair is not already included in the corpus: The agent would examine each pair by parsing the content of each one of the two webpages and comparing it with the content which is already indexed in the local database. If a ≥90% match is found for both pages, then the pair is discarded. If a <90% match is returned, the pair is eligible for further processing. Now if, for example, a pair returns a 99% match for one page and a 56% match for the other one, the pair would have to be kept, as this might be a case of a webpage whose translation was published before the proofreading, and then re-published with many changes in the content at a later time when the proofreading was completed.

III. Verification of parallel content: In order to verify that the pairs of webpages are mutual translations, the agent needs to identify and evaluate the similarities between them, at both structure and content levels. Some possible methods for doing this have been described above, as used by the three experimental tools. The main difficulty that complicates this action lies in the fact that the agent has to understand the text in order to be able to make content-based comparisons. But no effective text understanding can be envisaged without the proper identification and representation of the semantic content of the webpages. So, until the Semantic web becomes a reality, there are only a few ways available to tackle this problem. We can either use large bilingual lexica (the option followed by BITS and STRAND) or an index of domain-specific semantic keywords against which we must parse the paired webpages. These lexica or terminology lists can be the ones that the user has created and that reside in his terminology database.

IV. Download pairs of parallel webpages to a temporary local folder: This would allow the offline processing of the files, even if the user decides to interrupt the session at any point by going offline.

Offline actions

I. Non textual items filtering out: After having a validated set of parallel webpages, the agent would strip out any non textual elements (e.g. images, video, graphics, etc.) from every webpage

II. Conversion into plain Unicode text (with no indication of font or text formatting)

III. Tagging: Each text file would be tagged with the following information: date of creation of content, language, source (URL and webpage name, if it exists) and subject (if domain-specific keywords were used).

IV. Full-text indexing in the TM repository: The texts would be indexed as character strings and would be aligned according the segmentation rules and alignment methods followed by each TM system.

General functionality aspects

The agent would start operating once the TM application is opened and it would work in the background without interrupting the work of the translator. It would run in regular temporary sessions (e.g. once a week or month) either automatically (i.e. the translator would not have to initiate the process every time) or at the user's command. The module would be accessible through the main interface of the TM application and a separate dialog window would allow users to set the parameters under which the agent would operate. Such settings would include:

- Schedule of temporary sessions
- Determination of the size of the database (the agent would stop the session when a predefined size threshold was reached)
- Search restrictions: a) by website (the user would be able to restrict the search by listing specific websites that he considers more authoritative sources of good quality content; by default, the agent would search the entire public web), b) by domain (the user would be able to include subject-specific keywords in the search or import his bilingual glossary, if he wants to build a domain-specific parallel corpus).

This being a preliminary exploration of the possible integration of a corpus extraction tool in a TM system it leaves room for further investigation into all functionality aspects of such an integrated productivity system and the potential issues that may appear. The creation of a prototype would be the next step to test the new functionalities of the system and to make improvements to the model.

The integration of a corpus extraction facility in a TM system would mark an important development in translation technology. By offering the possibility of exploiting the potential of the web as the biggest electronic corpus, TM systems would become, undoubtedly, more powerful tools. Such a facility would enhance the functionality and expand the scope of use of existing TM systems, as well, it is believed, as extending the penetration of translation technology solutions among the large market share of translation professionals who are currently unconvinced of the benefits that technology can bring to their work.

7.3 The current and future role of TM systems

Einstein is often quoted as saying that his pencil was smarter than he was. By 'smartness' he was referring to his pencil's ability to fixate temporary results on paper, freeing up his working memory and thereby enabling him to solve an arithmetic problem in increments. Without the pencil and paper, he would have to store the results of the increments in short term memory, quickly exceeding its limits.

In the parallel universe of translation, the traditional role of Translation Memory systems has been that of a smart tool with unlimited memory which can free up the translator from menial tasks such as the repetitive translation of same text and let him concentrate on challenging translations. In their role as translation-support tools, they have also aimed at helping translation professionals to deal with complex translation projects and facilitating the collaboration across project members. Finally, their functions have aimed at allowing translators to build resources directly into their TM-enabled workflow, which can make a big difference to the consistency, quality, and speed of translation.

Several translator trainers have warned of the dangers of over-relying on TM systems and have expressed their fear that while TM systems become more sophisticated at suggesting translations, translators will use their imagination less and less [Bowker & Barlow, 2004:76-79]. The potential 'dumbing down' effect of TM systems derives from the observation that the more extensive the epistemic knowledge base a translator has at his disposal, the less time he has to spend applying heuristic strategies [Wilss,1996:57]. Consequently, continuous suggestions of ready-made translations may put to sleep many of his skills including his resourcefulness, creativity a nd curiosity, and might hamper his inventiveness when it comes to developing new problem solving methods. This fear resonates with American journalist and scholar Sydney J. Harris' frequently quoted concern which surrounds technology in general: "the real danger is not that computers will begin to think like men, but that men will begin to think like computers" [Eves, 2003:63].

The mitigation of the above risks cannot come from restricting the advancement of translation technology if there is a clear need calling for it. As Austermühl [2001:11] believes, "the effective use of translation technology starts from the translator's position.

The translator has to determine what types of technology are needed at what stages in the translation process in order to optimise his performance." Pym [2003: 494], on the same basis, stresses the importance of the cautious user when it comes to assessing the translations proposed by a TM system. He believes that translator training is responsible for the wise use of technology by translators, as they should be taught to mistrust and constantly evaluate their sources of information. He also thinks that trainers should encourage translators to be creative and not translate slavishly but to adapt and localise the text to suit the intended reader or user. And since TM systems – as they become more sophisticated – are now able to present a wider range of translation alternatives for one's source text in less time, the trainer must try to enhance the translator's ability to select only one viable target text from a series of possible translations, quickly and with justified confidence [Pym, 2003:494].

Hopefully, the current role of TM systems in the production of human translations seems to be clear for most translators familiar with this type of technology. As one of our survey respondents felt the need to clarify:

> *"The translation memory tool is only to assist for implementing the translation task in a faster manner, whereas the most important and biggest part of the translation is directly connected with the style and personal approach to the content of the translated document of the translator himself."*

However, as knowledge resources will grow larger and more accessible from or integrated into a TM system, we envision a new role for TM systems that will not only not pose a threat to traditional translators' skills, but will help translators develop them further. TM systems could become training tools in the hands of translator trainers and learning tools in the hands of translators. Corpora would enable the pedagogical function of TM systems.

Our anticipation draws upon the conclusions of a number of studies which have been carried out on the use of corpora for translator training [see Bowker & Barlow, 2004:74-5, Laviosa, 2003:112; Zanettin, 2002]. According to existing research, trainee translators with access to bilingual corpora show enhanced knowledge of the source language, clearer understanding of the subject matter and have better chances of identifying appropriate target language equivalents, during the course of their training programme [Laviosa, 2003:112]. Similarly, Danielson's research [2003] on monolingual corpora and their value for training translators has shown that translators who have access to large monolingual corpora use a wider variety of expressions and idiomatic patterns and feel comfortable with paraphrasing, as opposed to the general tendency of translators to get locked into the word choices and structures of the source text with the frequent result of using words that are out of fashion and more formal.

Evidently, with the introduction of corpora, a TM system could be repurposed to be used for training or professional development purposes, something that would surely expand the possibilities of TM technology.

Conclusion

The implementation of survey research proved to be largely successful in addressing the research questions of this investigation, in the sense that it delivered information of sufficiently high quality and detail to enable the generation of theories regarding translation professionals' needs. Operating across disciplines (Translation Studies, Social Studies, Requirements Engineering and Product Design) allowed us to employ multiple research methods and techniques during the analysis and interpretation of data, and showed that it is fruitful to follow an interdisciplinary approach to the problem of user needs identification.

This research has addressed the challenge of delving into the translator's world and revealed needs that are guided by practical experience and interests. It has established connections between needs and particular user groups, as well as contexts of use, and has provided design directions from the user's perspective on both functional and non-functional aspects of TM systems.

In summary, three key concepts corresponding to TM system properties appeared to bear the greatest importance to TM users, as they emerged again and again throughout the survey's responses: a TM system's *flexibility*, *simplicity* and *ease of access*. These three system properties, given that they are considered so important by users, are also in all probability the determining factors when selecting a TM system.

Flexibility seems to be attained when the system offers the possibility to the user of customising and defining the settings for most of its features. This way it can be adaptable to each user's personal needs. It is also achieved through the option of enabling or disabling certain features according to the different tasks performed by certain groups of users. Finally, a system is flexible when allowing the user to deactivate the tool for a specific part of the text where it is not required or deemed useful.

Simplicity is perceived in different ways by different user groups. For users with little or no experience in using TM systems and with an average level of IT skills, simplicity is the opposite of '*featurism*', meaning that the more features a TM system has, the more complicated it is to use. As a consequence, complexity in a software is more likely to discourage novice translators from using it. This group of users seem to prefer systems in 'turnkey' format which can be plugged in, turned on, and operated with little or no additional configuring. On the other hand, more experienced TM users with good or excellent IT skills favour more complex systems. For them, a feature-rich application does not necessarily mean an overbearing one, as long as the workflow that ties all features together is simple. Both user categories seem to agree on the idea that simplicity can be achieved by organising all features under the umbrella of a single simple process that supports the translation workflow from the beginning to the end. It is also believed that simplicity can be also achieved through a single software solution. According to TM Survey 2006 [Lagoudaki, 2006], the significant majority of respondents (75%) indicated a preference for a single TM software solution, instead of multiple applications that might be linked together. Similarly, based on the qualitative data of the survey, multiple applications – even if they are well linked to each other – are thought to be less user friendly and entail an additional learning curve.

Ease of access relates to the affordability of the system, not only in terms of purchase cost by also in terms of upgrade, support and training costs. According to TM Survey 2006 [Lagoudaki, 2006] cost is found to be the main reason for not using a TM system for 25% of the survey's respondents. In addition, our qualitative data revealed that affordable price is a persisting request, although some users appear willing to pay a premium price for tools with value-adding features.

A key finding of this research has also been the attitude of translation professionals towards automation. In general, TM users strongly believe that software should not be pushed to do anything unless it can do it both better and faster than a human. In other words, any technically possible enhancement, including the automation of particular tasks, will only be accepted by the users if it does not introduce new time-consuming tasks into the translation process.

The most frequently discussed tasks that users believed needed greater automation were the alignment of parallel texts, the analysis of project files and the detection of errors in a

translation through a user-defined quality control procedure. TM systems are expected to deliver satisfying results when performing these tasks with minimal or no supervision. Other tasks that involve the understanding of the meaning of words or texts, such as term extraction and translation assembly, would benefit from a degree of automation, but the user's input would be critical for the success of the task. Specifically, our research showed that the sophistication of the match assembly capabilities of a TM system, through the implementation of machine translation techniques, is more desirable by inexperienced translators. Experienced translators ardently believe that since creative rendering of meaning is still the domain of humans, a system's computational efforts must stop at the point where the imagination of the translator risks being compromised.

In view of the fact that current trends in translation technology are conspiring to automate the task of translation, developers of TM systems should never forget the intended use of these systems (which is different from that of an MT system), which is to support translators in their decision-making process and not try to automate this process by mechanically constructing language to be offered as a translation solution and requiring translators to validate the system's decisions. At the other end of the spectrum, translators, for their part, especially the inexperienced ones, should not over-rely on the MT capabilities of a TM system, as disappointing results can create a contemptuous attitude towards the technology which under certain circumstances might prove pain-saving. The key to a mutually beneficial relationship between a translator and an MT feature appears to be the cautious application of MT and the consultation of MT suggestions as if they were just another source of reference.

Overall, the survey revealed a large number of needs that are not yet met by most if not all of the existing TM systems. Some respondents formed their requests out of their current problems. Others visualised future TM systems by imagining what would make their job easier.

Some of the ideas and suggestions in this thesis have already found their way into improved versions of TM systems through regular dissemination of the present research's findings during public talks and presentations. As regards future developments, it is envisaged that as translation professionals become savvier about IT, their expectations from technology will rise. Their expectations will also expand with every successful software product they see, and consequently their needs will become more sophisticated, calling for an array of additional intelligent features. What is more certain than any other prediction for the future of translator-support technology is that all advances will derive from the language services industry's needs.

We can only see a short distance ahead, but we can see plenty there that needs to be done. We anticipate hybrid TM platforms that integrate a number of applications, techniques and resources. By looking at other office applications and their evolution through time, we also

anticipate convergence, that is, providing translation professionals with the ability to access multiple and convenient tools via multiple interfaces at the time and in the format they require. A system of value-added services is also likely to be developed around TM products. However, for services and tools to be constantly evolving and adapting to the growing needs of their users, better understanding of the users themselves will be key.

We also foresee a shift in the pricing model for TM products, as R&D investment is likely to turn to the development of language resources and to tools that can access, manage and maintain such resources. The price of technology will drop significantly (with core TM systems made available at little or no cost), whereas add-ins such as application plug-ins and resources will be sold separately. Such a shift seems to make sense, considering the diversity of user needs and the fact that a single TM system, no matter how comprehensive it is in terms of functionality, cannot satisfy all its users to the same degree.

Finally, perhaps some day in the distant future Christopher Evans' [1979] foresight will come true even for TM systems: computer engineers will develop means for computers to programme themselves, so a translator will be able to specify his requirements and the tool will seamlessly design and create a program that will be custom-made for her needs.

Until that day comes, software engineers are responsible for producing software that is simple and elegant, but also moral and ethical.

Indeed, software engineers have a great impact on the translation professional's work experience, as software can enhance their productivity and amplify human intelligence. However, they often live in a slightly different world from the practitioners, i.e. translation professionals. And if they have come up with some good features – as they all have – they tend to focus on them instead of going on a truly never-ending quest for the ever-more-sharpened tool.

The commonest danger is to design software based on the development team's intuition and to make assumptions concerning the translators' needs. The difference between what software engineers think is important and what really is important or valuable creates waste for the developers [Beck and Andres, 2004:17], while the resulting software can be so user-hostile that it can damage the confidence of translators in such tools. It is therefore recommended that user engagement be pursued in all stages of software development.

First and foremost, quality should be built into the software in the first stages of a development project, working in conjunction with end-users. When undertaking the development and implementation of such systems it is important that developers consider how the systems will actually function in coordination with human beings in a translation workflow process. By observing how people use technology now, greater levels of product usability can be achieved. A steady flow of early, concrete and continuous feedback will also help to ensure the compliance of the system's design with users' requirements. In fact, the

sooner the designer has domain-specific knowledge, the sooner she can adapt the design. Refining the design of the system must be straightforward. The actual design will never be a perfect reflection of the ideal, but TM software designers need to strive daily to bring the two closer.

Users should be involved in software testing, as well. They should be allowed to monitor the progress of development, which will in turn allow the system to evolve in the right direction and to catch defects early.

In addition to the above, engineers must incorporate change into the development process. Everything in software must be adaptable, since the requirements change, the translation business changes, the technology changes and the translator's profession changes. Constant feedback from the users is again the key to developing software that is in tune with the ever-changing modern business world. Similarly, a keen eye on the translator's work reality will help software engineers react to changing customer needs, anticipate market changes and keep up to date with developments in the industry and technology.

Software engineers generally seem to love complexity. But the more complicated the software gets, the greater the need to focus not only on quality assurance testing but also on usability testing points that keep learning curve and human factors in mind. If a tool becomes so complicated that a long (and expensive) training course is needed, then users risk opting out of the training, and then struggling for a long time trying on their own to be productive with the new tool. For these reasons, software engineers need to take a step back more often and simplify things. In fact, depending on the target user groups, the way forward for some systems could be not sophistication, but simplification.

Finally, engineers must show respect for the user whose life may be affected by the software, and bear in mind that their passion for creating software has the potential to improve the quality of work and life of a translation professional.

The goal is no longer simply to meet specifications, it is to deliver software so superb that user expectations are not just met, but exceeded. In other words, users should not be simply satisfied, but delighted. Software developers must make sure they address expected requirements, performance requirements, as well as exciting requirements that will delight users when delivered. Inspections and testing can be the answer to defect reduction, but they cannot increase value. To deliver value, developers should not study defects; they must study the needs of their software's users.

Functionality, attractiveness, ease of use, affordability, and portability are taken for granted. The new trends are for applications that inspire users, enhance their lives and evoke emotions and dreams. This requires research into conscious as well as unconscious processes that are related to positive emotions and pleasurable experiences. A well-

structured research effort will open up a field for human-centred, innovative product and service developments and help to minimise risks in the research and development process.

In technology nothing is impossible. Research in software engineering and natural language processing is evolving at a rapid pace, surmounting programming limitations of the past. The question at the moment is not whether it would be possible to develop some features, but whether they would bring any real benefit to the daily work of the translation professional and therefore would justify the development effort. This is something that TM developers need to find out through a careful and systematic requirements engineering process, carried out in close cooperation with the end-users of their products.

Translation professionals, for their part, must keep an open mind about technology and grasp any opportunity for familiarising themselves with it and testing it in the context of their work. Only by exploring the possibilities of technology out of personal interest or curiosity can one actually appreciate its real purpose. In addition, those interested in technology must seek collaboration with translator-support tool developers and offer feedback through any of the available channels of communication, in order to improve existing systems. Finally, they must realise that technology presents a choice and that people ought to use what they need from it. It is not meant to lead people, but rather people should lead it to serve their own personal goals.

References

1. **Agouridas, Vassilis, Henri Winand, Alison McKay and Alan de Pennington** (2008) "Advanced product planning: a comprehensive process for systemic definition of new product requirements". *Journal of Requirements Engineering*, Vol. 13, No. 1, pp. 19-48.
2. **Allan, George** (2003) "A critique of using Grounded Theory as a research method". *EJBRM,* Vol. 2, Issue 1.
3. **Arthern, Peter** (1979) "Machine translation and computerised terminology systems: a translator's viewpoint." In Snell, Barbara (ed.) *Translating and the computer* [Proceedings of a seminar held in London (14 November 1978).] Amsterdam: North-Holland, pp. 77-108.
4. **Asensio, Roberto Mayoral** (2003) *Translating Official Documents*. Manchester: St. Jerome Publishing.
5. **Austermühl, Frank** (2001) *Electronic Tools for Translators*. Manchester: St. Jerome Publishing.
6. **Babych, Bogdan, Anthony Hartley and Serge Sharoff** (2007) "A Dynamic Dictionary for Discovering Indirect Translation Equivalents". In proceedings of the International Conference *Translating and the Computer 29*, 29-30 November 2007, London. London: Aslib.
7. **Baker, Mona** (1992) *In other words*. London & New York: Routledge.
8. **Basit, Tehmina** (2003) "Manual or electronic? The role of coding in qualitative data analysis". In *Educational Research*, Vol. 45, No. 2, summer 2003, pp. 143-154.
9. **Battle, Lisa** (2005) "Patterns of Integration: Bringing User Centered Design into the Software Development Lifecycle". In *Human-Centered Software Engineering — Integrating Usability in the Software Development Lifecycle*. Netherlands: Springer, pp. 287-308.
10. **Beck, Kent and Cynthia Andres** (2004) *Extreme Programming Explained*. Boston: Addison-Wesley.
11. **Benis, Michael** (2007a) "Déjà Vu: taking a second look at CAT in a mature market". *ITI bulletin*, November-December 2007.
12. **Benis, Michael** (2007b) "Searching for solutions". *ITI bulletin*, January-February 2007.
13. **Benis, Michael** (2005) "SDLX™: The dark horse of translation memory." *ITI Bulletin*, Vol. 22, No. 27, March-April 05.
14. **Benis, Michael** (2003a) "Much more than memories". *ITI bulletin*, November-December 2003. Available at http://www.atril.com/docs/Benis-ITI-DVX.pdf - last accessed on 17/09/2008)
15. **Benis, Michael** (2003b) "Letting the CAT out of the bag". *The Linguist*, Vol. 42, No. 3, June-July 2003.
16. **Benis, Michael** (2000) "How the memory measured up". Article available on the website of TransRef (URL: http://www.transref.org/default.asp?docsrc=/u-articles/Benis4.asp - last accessed on 07/09/2008)
17. **Bey, Youcef, C. Boitet and K. Kageura** (2006) "The TRANSBey Prototype: An Online Collaborative Wiki-Based CAT Environment for Volunteer Translators". Yuste, E., (Ed.). In the Proceedings of the 3rd *International Workshop on Language Resources for Translation Work, Research & Training (LR4Trans-III)*. LREC 2006 - Fifth International Conference on Language Resources and Evaluation. 28th May 2006. Magazzini del Cotone Conference Centre, Genoa, Italy. Paris: ELRA / ELDA (European Language Resources Association, European Language Resources Distribution Association), pp. 49-54.
18. **Biau Gil, José Ramón** (2007) "What You See is What You Get? A Pilot Experiment on Access to Visual Information in Translation Interfaces.". Presentation at the 12th Annual Internationalisation and Localisation Conference, 27 September 2007, Dublin, Ireland.
19. **Biau Gil, José Ramón** (2005) "'Please translate close to the source text, i.e. deliver a good translation' – Trust and Technology Problems Related to Literal or Free Translation". In *Génesis. Revista Científica do ISAI.* nº 5, pp. 77-82.

20. **Bowker, Lynne** (2005) "Productivity vs. Quality? A Pilot Study on the Impact of Translation Memory Systems". *Localisation Focus* 4 (1), pp. 13-20.
21. **Bowker, Lynne and Michael Barlow** (2004) "Bilingual concordancers and translation memories: A comparative evaluation." In proceedings of the 20th International Conference of Computational Linguistics COLING-2004, Geneva, Switzerland. Second International Workshop on Language Resources for Translation Work, Research & Training, pp. 70-83.
22. **Bowker, Lynne** (2002) *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
23. **Callison-Burch, Chris, Colin Bannard and Josh Schroeder** (2004) "Searchable Translation Memories". In proceedings of the International Conference *Translating and the Computer 26*, 18-19 November 2004, London. London: Aslib.
24. **Carl, Michael, Andy Way, Reinhard Schäler** (2002) "Toward a Hybrid Integrated Translation Environment". In proceedings of the 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002, 8-12 October 2002, Tiburon, CA, USA.
25. **Carl, Michael** (2000) "Combining Invertible Example-Based Machine Translation with Translation Memory Technology". In proceedings of the *4th Conference of the Association for Machine Translation in the Americas, AMTA 2000: Envisioning Machine Translation in the Information Future*, 10-14 October 2000, Cuernavaca, Mexico.
26. **Carl, Michael and Silvia Hansen** (1999) "Linking Translation Memories with Example-Based Machine Translation". In O. Streiter, M. Carl & J. Haller (eds.) *Hybrid Approaches to Machine Translation*. WP36/99. Saarbrücken: IAI, 1999.
27. **Carter-Sigglow, Janet** (2004) "Who's afraid of CAT? Redefining the boundaries of translation." In proceedings of the International Conference *Translating and the Computer 26*, 18-19 November 2004, London. London: Aslib.
28. **Champollion, Yves** (2003) "Convergence in CAT: blending MT, TM, OCR & SR to boost productivity". In proceedings of the International Conference *Translating and the Computer 25*, 20-21 November 2003, London. London: Aslib.
29. **Chen, Jiang and Jian-Yun Nie** (2000) "Parallel Web text mining for cross-language information retrieval". In proceedings of *Recherche d'Informations Assistée par Ordinateur (RIAO)*, 12-14 April 2000, Paris, pp. 62–77.
30. **Chenon, Cristophe** (2004) "Dynamical visualisation of nested correspondences". In proceedings of the International Conference *Translating and the Computer 26*, 18-19 November 2004, London. London: Aslib.
31. **Civera, Jorge, Antonio L. Lagarda, Elsa Cubel, Francisco Casacuberta, Enrique Vidal, Juan M. Vilar, & Sergio Barrachina** (2006) "A computer-assisted translation tool based on finite-state technology". In proceedings of *EAMT-2006: 11th Annual Conference of the European Association for Machine Translation*, June 19-20, 2006, Oslo, Norway, pp.33-40.
32. **Couper, Mick P.** (2000) "Web surveys: A review of issues and approaches". *Public Opinion Quarterly* 6(4), pp. 464-494.
33. **Cronholm, Stefan** (2002) "Grounded Theory in Use – A Review of Experiences". In the proceedings of the European Conference on Research Methodology for Business and Management Studies, Reading University, Reading, UK.
34. **Danielsson, Pernilla** (2003) "Units of Meaning in Translation – How to Make Real Use of Corpus Evidence". In proceedings of the International Conference *Translating and the Computer 25*, 20-21 November 2003, London. London: Aslib.

35. **Davies, Alan** (1993) *Software Requirements: Objects, Functions and States*. Englewood Cliffs, New Jersey: Prentice-Hall.
36. **Désilets, Alain** (2008) "WeBiText: building Large Heterogeneous Translation Memories from Parallel Web Content". In proceedings of the International Conference *Translating and the Computer 30*, 27-28 November 2008, London. London: Aslib.
37. **Désilets, Alain, Lucas Gonzalez, Sébastien Paquet, Marta Stojanovic** (2006) "Translation the Wiki Way". In the proceedings of the *2006 International Symposium on Wikis*. ACM Press, pp. 19-32.
38. **Dillman, Don** (2007) *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.
39. **Dillon, Sarah and Janet Fraser** (2007) "Translators and TM: An investigation of translators' perceptions of translation memory adoption". *Machine Translation* journal, Springer Netherlands. Available online from SpringerLink (www.springerlink.com).
40. **Drugan, Joanna** (2004) "Multilingual document management and workflow in the European institutions". In proceedings of the International Conference *Translating and the Computer 26*, 18-19 November 2004, London. London: Aslib.
41. **Eberlein, Armin and Julio Cesar Sampaio do Prado Leite** (2002) "Agile Requirements Definition: A View from Requirements Engineering". In proceedings of the International Workshop on Time Constrained Requirements Engineering, Essen, September 2002.
42. **Esselink, Bert** (2003) "Localisation and translation". In *Computers and Translation - A translator's guide* (ed. Harold Somers). Amsterdam and Philadelphia: John Benjamins Publishing.
43. **Esselink, Bert** (2000) *A Practical Guide to Localisation*. Amsterdam and Philadelphia: John Benjamins Publishing.
44. **European Commission LRE** (1997) "TEMAA Project: Final Report". October 1997.
45. **European Commission, Translation DG** (2007) "Translation Tools and Workflow". Directorate-General for Translation of the European Commission, February 2007.
46. **Evans, Christopher Riche** (1979) *The mighty micro : the impact of the computer revolution*. London: Gollancz.
47. **Eves, Howard W.** (2003) Mathematical Circles: Volume III, Mathematical Circles Adieu & Return to Mathematical Circles, The Mathematical Association of America, Inc., p. 63.
48. **Farghaly, Ali** (ed.) (2003) *Handbook for Language Engineers*. Leland Stanford Junior University: CSLI Publications.
49. **Fernández, Walter D.** (2005) in Dennis Hart and Shirley Gregor (eds) *Information Systems Foundations: Constructing and Criticizing*. Canberra: Australian National University E-Press, pp. 43-59.
50. **Fulford, Heather and Joaquín Granell Zafra** (2004) "The uptake of online tools and web-based language resources by freelance translators: implications for translator training, professional development, and research". In proceedings of the *II International Workshop on Language Resources for Translation Work, Research and Training*, COLING 2004, Geneva.
51. **Gale, William and Kenneth Church** (1993) "A program for aligning sentences in bilingual corpora". *Computational Linguistics*, Vol. 19, issue 1, pp. 75-102.
52. **Gamallo Otero, Pablo** (2007) "Learning bilingual lexicons from comparable English and Spanish corpora". In the proceedings of *MT Summit XI*, 10-14 September 2007, Copenhagen, Denmark, pp.191-197
53. **Garcia, Ignacio and Vivian Stevenson** (2006) "Lingotek: review". *MultiLingual Computing*, No. 84, Vol. 17, issue 8. December 2006.
54. **Gervais, Daniel** (2002) "The Full-Text Multilingual Corpus: Breaking the Translation Memory Bottleneck" MultiCorpora R&D Inc. White Paper, April 2002.

55. **Giles, Jim** (2005) "Internet encyclopaedias go head to head". *Nature*, December 2005.
56. **Gilgun, Jane F.** (2007) "The Analysis of Qualitative Data". Available at http://ssw.cehd.umn.edu/img/assets/5661/Notes_Analysis_Qualitative_Data_Rev0307.pdf [last accessed on 7/3/2008]
57. **Glaser, Barney** (1998) *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA: Sociology Press.
58. **Glaser, Barney** (1992) *Emergence vs. Forcing: Basics of Grounded Theory Analysis*. Mill Valley, CA: Sociology Press.
59. **Glaser, Barney and Anselm Strauss** (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Co.
60. **Godden, Kurt** (2000) "J2450 Translation Quality Metric". *Society of Automotive Engineers*, March 2000.
61. **Gotel, Orlena C. Z. and Anthony C. W. Finkelstein** (1994) "An Analysis of the Requirements Traceability Problem". In proceedings of the International Conference on Requirements Engineering, Colorado Springs, 1994.
62. **Gotti, Fabrizio, Philippe Langlais, Elliott Macklovitch, Didier Bourigault, Benoit Robichaud and Claude Coulombe** (2005) "3GTM: A Third-Generation Translation Memory". In proceedings of the *3rd Computational Linguistics in the North-East (CLiNE) Workshop*, Gatineau, Québec, 26 August 2005, p. 8.
63. **Goto, Isao, Naoto Kato, Noriyoshi Uratani, Terumasa Ehara, Tadashi Kumano and Hideki Tanaka** (2003) "A Multi-language Translation Example Browser". In proceedings of the *Machine Translation Summit IX*, 23-27 September 2003, New Orleans, Louisiana, USA.
64. **Goto, Isao, Naoto Kato and Terumasa Ehara** (2001) "A multilingual news database and its application to a translation memory system". In the proceedings of the *6th Natural Language Processing Pacific Rim Symposium*, Workshop on Language Resources, 30 November 2001, Tokyo, Japan.
65. **Gough, Nano, Andy Way, Mary Hearne** (2002) "Example-Based Machine Translation via the Web". In proceedings of the 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002 Tiburon, CA, USA, October 8 – 12, 2002.
66. **Gow, Francie** (2004) "Extracting Useful Information from TM Databases". *Localisation Reader* 2004-2005, pp. 41-4.
67. **Gow, Francie** (2003) *Metrics for Evaluating Translation Memory Software*. MA Thesis. University of Ottawa, Canada.
68. **Griffin, Abbie and John R. Hauser** (1993) "The Voice of the Customer". *Marketing Science*, Vol. 12, No. 1, Winter 1993, pp. 1-27.
69. **Grönroos, Mickel and Ari Becks** (2005) "Bringing Intelligence to Translation Memory Technology". In proceedings of the International Conference *Translating and the Computer 27*, 24-25 November 2005, London. London: Aslib.
70. **Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer and Roger Tourangeau** (2004) Survey Methodology. New Jersey: John Wiley & Sons.
71. **Ha, Lee An, Fernandez, Gabriela, Mitkov, Ruslan and Corpas, Gloria** (2008a) "Mutual terminology extraction using a statistical framework". In the proceedings of *SEPLN 2008*, Madrid, Spain, pp. 107-112.
72. **Ha, Lee An, Fernandez, Gabriela, Mitkov, Ruslan and Corpas, Gloria** (2008b) "Mutual bilingual terminology extraction". In proceedings of the Sixth International Language Resources and Evaluation (*LREC'08)*, Marrakesh, Morocco.
73. **Halpern, Jack** (2004) "Issues in Japanese Information Retrieval". *Localisation Reader* 2004-2005.

74. **Harris, Brian** (1988) "Bi-text, a New Concept in Translation Theory". In *Language Monthly*, No. 54, pp.8-10.
75. **Hatim, Basil and Jeremy Munday** (2006) *Translation: An advanced resource book*. Oxon/New York: Routledge.
76. **Herzwurm, Georg, Sixten Schockert and Werner Mellis** (1999) "Higher customer satisfaction with prioritizing and focused Quality Function Deployment". In proceedings of the sixth European Conference on Software Quality, Vienna, Austria. Available at: http://www.bwi.uni-stuttgart.de/fileadmin/abt9/Publikationen_Herzwurm3/paper.pdf.
77. **Hodász, Gábor, Tamás Gröbler, & Balázs Kis** (2004) "Translation memory as a robust example-based translation system". In proceedings of the *9th EAMT Workshop, "Broadening horizons of machine translation and its applications"*, 26-27 April 2004, Malta, pp.82-89.
78. **Hodász, Gábor & Gábor Pohl** (2005) "MetMorpho TM: a linguistically enriched translation memory". In proceedings of the International workshop: *Modern approaches in translation technologies,* Borovets, Bulgaria, 24 September 2005, pp. 26-30.
79. **Höge, Monika** (2002) *Towards a Framework for the Evaluation of Translators' Aids Systems*. PhD thesis. Helsinki University, Helsinki.
80. **Holmes, James S.** (2000) "The Name and Nature of Translation Studies". In Lawrence Venuti (ed.) (2000) The Translation Studies Reader. London: Routledge; pp.172-85. [This book chapter is an expanded version of a paper presented in the Translation Section of the Third International Congress of Applied Linguistics, held in Copenhagen, 21-26 August 1972.]
81. **Howe, Jeff** (2008) *Crowdsourcing. Why the Power of the Crowd Is Driving the Future of Business*. New York: Crown Business.
82. **Hoye, Leo Francis and Louisa S.W. Chan** (2007) "Pragmatics and multimodal discourse: The verbal-visual interface". In proceedings of the *X Symposium on Social Communication*, 22-26 January 2007, Santiago de Cuba, pp. 253-6.
83. **Huang, Jin-Xia, Wei Wang and Ming Zhou** (2003) "A Unified Statistical Model for Generalized Translation Memory System". In the proceedings of *MT Summit IX*, 23-27 September 2003, New Orleans, Louisiana, USA.
84. **Hunt, Timothy** (2006) "Will Language Search Engines replace Translation Memory Systems?" Article available on the Lingotek website at: http://www.lingotek.com/lsereplacetm.pdf.
85. **Hunt, Timothy** (2003) "Translation Technology Failures and Future". *Globalisation Insider*, Vol. XII, issue 1.4. Available at http://www.translationdirectory.com/article440.htm - last accessed on 07/10/2005.
86. **Hutchins, John** (2004) "Current commercial machine translation systems and computer-based translation tools: system types and their uses". Unpublished article available at http://ourworld.compuserve.com/homepages/wjhutchins/IJT-2004.pdf - last accessed on 07/10/2005.
87. **Hutchins, John** (2003) "Machine Translation: General Overview". In Ruslan Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press, pp. 501-11.
88. **Hutchins, John** (2001) "Machine translation and human translation: in competition or in complementation?" *International Journal of Translation*, Vol.13, No.1-2, Jan-Dec 2001, pp. 5-20.
89. **Hutchins, John** (1998) "The origins of the translator's workstation" *Machine Translation* 13 (4), pp. 287-307.
90. **Hutchins, John and Harold Somers** (1992) *An Introduction to Machine Translation*. London: Academic Press Limited.
91. **Isabelle, Pierre** (1993) "Machine-Aided Human Translation and the Paradigm Shift". In the proceedings of *Machine Translation Summit IV*, Kobe, Japan, 1993.

92. **ISO/IEC 9126** (Standard): Information technology - Software Product Evaluation - Quality characteristics and guidelines for their use. International Organisation for Standardisation, 1991.
93. **Iverson, Steve** (2004) "Working with Translation Memory". *Localisation Reader* 2004-2005.
94. **Jackson, Michael** (1995) *Requirements and Specifications*. Wokingham: Addison Wesley.
95. **Järvelin, Kalervo and Ari Pirkola** (2005) "Morphological Processing in Mono- and Cross-Lingual Information Retrieval". *Inquiries into Words, Constraints and Contexts*, Stanford University: CSLI Publications, pp. 214-226.
96. **Joy, Lorna** (2002) "Translating tagged text – imperfect matches and a good finished job". In proceedings of the International Conference *Translating and the Computer 24*, 21-22 November 2002, London. London: Aslib.
97. **Kamps, Jaap, Christof Monz, Maarten de Rijke and Börkur Sigurbjörnsson** (2003) "Language-dependent and Language-independent Approaches to Cross-Lingual Text Retrieval". In proceedings of 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 21-22, 2003, pp. 152-165.
98. **Kano, Noriaki** (1984) "Attractive quality and must-be quality". The Journal of the Japanese Society for Quality Control, April, pp. 39-48.
99. **Kay, Martin** (1980/1997) "The Proper Place of Men and Machines in Language Translation." *Machine Translation*, Vol. 12, numbers 1-2, March 1997, pp. 3-23.
100. **Kendall, Jason** (1999) "Axial Coding and the Grounded Theory Controversy". *Western Journal of Nursing Research*, Vol. 21, issue 6.
101. **Kim, Ryonhee** (2006) "Use of Extralinguistic Knowledge in Translation". META, Vol. 51, No. 2, June 2006, pp. 284-299.
102. **King, Margaret** (1997) "Evaluation Design: The EAGLES framework" Konvens 97.
103. **Kingscott, Geoffrey** (2002) "Technical Translation and Related Disciplines". In: Perspectives: Studies in Translatology, Vol. 10:4. pp. 247-255.
104. **Kohler, Kirstin and Barbara Paech** (2002) "Requirement Documents that Win the Race: Not Overweight or Emaciated but Powerful and in Shape". In proceedings of the International Workshop on Time Constrained Requirements Engineering, Essen, September 2002.
105. **Kotonya, Gerald and Ian Sommerville** (1998) *Requirements Engineering*. Chichester: John Wiley & Sons.
106. **Kovitz, Ben** (2003) "Hidden Skills that Support Phased and Agile Requirements Engineering". *Requirements Engineering Journal*, Vol. 8, 2003.
107. **Kraaij, Wessel, Jian-Yun Nie, Michel Simard** (2003) "Embedding web-based statistical translation models in cross-language information retrieval". *Computational Linguistics*, Volume 29, Issue 3, pp. 381-419.
108. **Kreckwitz, Stefan** (2007) "Future Web-Based Translation Environments". In the proceedings of the 12th Annual Internationalisation and Localisation Conference, 26-28 September 2007, Dublin, Ireland.
109. **Krogstie, John** (1999) "Using Quality Function Deployment in Software Requirements Specification", In proceedings of the Fifth International Workshop on Requirements Engineering: Foundations for Software Quality (REFSQ'99), Heidelberg, Germany 14-15 June, pp. 171-187,
110. **Kussmaul, Paul** (2005) "Translation through Visualization". META, Vol. 50, No. 2, April 2005, pp. 378-391.
111. **Lagoudaki, Elina** (2006) "Translation Memories Survey 2006: Users' perceptions around TM use". In proceedings of the ASLIB International Conference *Translating & the Computer 28*, 16-17 November 2006, London, UK.

112. **Langlais, Philippe and Michel Simard** (2001) "Récupération de segments sous-phrastiques dans une mémoire de traduction". In proceedings of the 8th Conference *Traitement Automatique des Langues Naturelles (TALN)*, Tours, France.
113. **Laviosa, Sara** (2003) "Corpora and the translator". In Harold Somers (ed.) *Computers and Translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing, pp. 105-117.
114. **Laviosa, Sara** (ed.) (1998) "The Corpus-based Approach: A New Paradigm in Translation Studies". Special issue of *Meta* 43(4).
115. **Lengyel, István** (2006) "Controlling the workflow in translation projects". *MultiLingual Computing*, No. 84, Vol. 17, issue 8. December 2006.
116. **Levitt, Garry** (2004) "New Developments in TM Sharing". *MultiLingual Computing*. No. 67, Vol. 15, issue 7.
117. **Lewis, Barry** (2004) "NVivo 2.0 and ATLAS.ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs". *Field Methods*, Vol. 16, No. 4, November 2004, pp. 439–469.
118. **Lippmann, Erhard** (1971) "An Approach to Computer-Aided Translation." *IEEE Transactions on Engineering Writing and Speech*, Vol. 14, pp. 10-33.
119. **Lommel, Arle** (2004) "LISA 2004 Translation Memory Survey". Available to download from LISA's website (URL: http://www.lisa.org - last accessed on 17/10/2005).
120. **Lörscher, Wolfgang** (1996) "A Psycholinguistic Analysis of Translation Processes", Meta 41(1): 26-32.
121. **Ma, Xiaoyi and Mark Liberman** (1999) "Bits: A method for bilingual text search over the Web". In proceedings of *Machine Translation Summit VII*, 13-17 September 1999, Singapore.
122. **Macklovitch, Elliott** (2000) "Two Types of Translation Memory". In proceedings of the International Conference *Translating and the Computer 22*, 16-17 November 2000, London. London: Aslib.
123. **Macklovitch, Elliot and G. Russell** (2000) "What's been forgotten in Translation Memory". In the proceedings of the *4th Conference of the Association for Machine Translation in the Americas (AMTA 2000)*, 10-14 October 2000, Cuernavaca, Mexico, pp.137-146.
124. **Macklovitch, Elliott, Michel Simard and Philippe Langlais** (2000) "TransSearch: A Free Translation Memory on the World Wide Web". In the proceedings of LREC 2000, 2nd International Conference on Language Resources & Evaluation.
125. **Maiden. N.A.M, & Rugg. G** (1996) "ACRE: selecting methods for requirements engineering", software engineering journal, Vol. 2:3, pp. 183-192.
126. **Makoushina, Julia** (2007) "Translation Quality Assurance Tools: Current State and Future Approaches". In proceedings of the International Conference *Translating and the Computer 29*, 29-30 November 2007, London. London: Aslib.
127. **McCallum, Bob** (2004) "Translation Technology at the United Nations". *Localisation Reader* 2004-2005.
128. **Melby, Alan** (1995) "Why can't a computer translate more like a person?" Lecture delivered at the Brigham Young University, USA.
129. **Melby, Alan** (1992) The Translator Workstation, In *Computers in Translation: A Practical Appraisal*, (Ed. Newton, J.) Routledge, London, pp. 147-165.
130. **Melby, Alan** (1982) Multi-level translation aids in a distributed system In *Proceedings of COLING 82*, (Ed. Horecký, J.) North Holland Publishing Company, Amsterdam, pp. 215-220.
131. **Menezes, Arul and Stephen Richardson** (2001) "A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora". In proceedings of *MT Summit VIII*, Santiago De Compostela, Spain.
132. **Mikheev, Andrei** (2003) "Text segmentation". In Ruslan Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press, pp. 201-218.

133. **Miles, M. and M. Huberman** (1994) *Qualitative Data Analysis: An expanded source book?* Thousand Oaks, CA: Sage publications.
134. **Mitkov, Ruslan and Gloria Corpas** (2008) "Improving Third Generation Translation Memory systems through identification of rhetorical predicates". In proceedings of the *LangTech 2008* conference, 28-29 February 2008, Rome.
135. **Murata, Toshiki, Mihoko Kitamura, Tsuyoshi Fukui, and Tatsuya Sukehiro** (2003) "Implementation of collaborative translation environment 'Yakushite Net'". In the proceedings of MT Summit IX, pp. 479-482.
136. **Musale, Shailendra** (2004) "Getting More From Translation Memories". *Localisation Reader* 2004-2005.
137. **Nirenburg, Sergei, Jaime Carbonell, Masaru Tomita and Kenneth Goodman** (1992) *Machine Translation: A Knowledge-Based Approach*. San Mateo, California: Morgan Kaufmann Publishers.
138. **Norman, Donald A.** (1986) "Cognitive engineering". In: Norman, Donald A., Draper, Steven (Eds.) *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 31-61.
139. **O'Brien, Sharon** (2006) "Eye-tracking and Translation Memory Matches". Perspectives: Studies in Translatology, Vol. 14, No. 3, pp.185-205.
140. **O'Hagan, Minako and David Ashworth** (2002) *Translation-mediated Communication in a Digital World: Facing the Challenges of Globalization and Localization (Topics in Translation S.)*. London: Multilingual Matters.
141. **O'Leary, Tara** (2000) "Memory Based Translation: Using Translation Memory Tools, from a Freelancer's Perspective". In proceedings of the International Conference *Translating and the Computer 22*, 16-17 November 2000, London. London: Aslib.
142. **Oliver, Antoni and Mercè Vàzquez** (2007) "A Free Terminology Extraction Suite". In proceedings of the International Conference *Translating and the Computer 29*, 29-30 November 2007, London. London: Aslib.
143. **Ornstein, Michael** (1998) "Survey research". *Current Sociology* 46(4), pp. 1-143.
144. **Pekar, Victor and Ruslan Mitkov** (2007) "New Generation Translation Memory: Content-Sensivite Matching". In proceedings of the 40th Anniversary Congress of the Swiss Association of Translators, Terminologists and Interpreters, 29-30 September 2006, Bern (ASTTI, 2007).
145. **Planas, Emmanuel** (2005) "SIMILIS: Second-generation translation memory software". In proceedings of the International Conference *Translating and the Computer 27*, 24-25 November 2005, London. London: Aslib.
146. **Pym, Anthony** (2006) "Translation technology as rupture in the philosophy of dialogue" In proceedings of the sixth Portsmouth Translation Conference on *Translation Technologies and Cultures*, October 2006.
147. **Pym, Anthony** (2003) "Translational ethics and electronic technologies". In Profissionalização do Tradutor, Lisboa: Fundação para a Ciência e a Tecnologia / Uniao Latina, 2004, pp. 121-126.
148. **Quah, Chiew Kin** (2006) *Translation and Technology*. Basingtoke/New York: Palgrave Macmillan.
149. **Ramesh, Bala and Matthias Jarke** (1999) "Towards Reference Models for Requirements Traceability". *IEEE. Transactions on Software Engineering*, Vol. 27, No. 1, pp. 58-93
150. **Ray, Deborah S. and Eric J. Ray** (1999) "Good, Fast, Cheap: Translation Memory Systems Offer the Potential for All Three" Technical Communication, V.46, No. 2, May 1999.
151. **Reynaert, Martin** (2005) *Text-Induced Spelling Correction*. PhD thesis. Tilburg University, The Netherlands.

152. **Reinke, Uwe** (1999) "Towards a Closer Integration of Termbases, Translation Memories, and Parallel Corpora: A Translation-Oriented View". In: Sandrini, Peter (Hrsg.): TKE '99 - Terminology and Knowledge Engineering, Innsbruck, 23.-27. August 1999. Wien: TermNet, pp. 527-543.
153. **Resnik, Philip and Noah A. Smith** (2003) "The Web as a parallel corpus" *Computational Linguistics*, Volume 29 , Issue 3, pp. 349-380.
154. **Resnik, Philip** (1998) "Parallel strands: A preliminary investigation into mining the Web for bilingual text". In proceedings of the *Third Conference of the Association for Machine Translation in the Americas (AMTA-98)*. 28-31 October 1998, Langhorne, PA.
155. **Rowlands, Bruce** (2005) "Grounded in Practice: Using Interpretive Research to Build Theory". *The Electronic Journal of Business Research Methodology*, Vol. 3, Issue 1, pp. 81-92.
156. **Ryan, Andrew J.** (2000) "An Approach to Quantitative Non-Functional Requirements in Software Development" In proceedings of the 34th Annual Government Electronics and Information Association Conference, 2000 Dallas, Texas.
157. **Sager, Juan C.** (1994) *Language Engineering and Translation. Consequences of automation*. Amsterdam: Benjamins.
158. **Samuelson-Brown, Geoffrey** (2004) *A Practical Guide for Translators*. Fourth revised edition. London: Multilingual Matters.
159. **Saralegi, Xabier, Iñaki San Vicente and Maddalen López de Lacalle** (2008) "Mining Term Translations from Domain Restricted Comparable Corpora". In the proceedings of *SEPLN 2008*, Madrid, Spain, pp. 273-281.
160. **Sarker, Suprateek, Francis Lau and Sundeep Sahay** (2001) "Using an Adapted Grounded Theory Approach For Inductive Theory Building About Virtual Team Development". *Database for Advances in Information Systems*, Vol. 32, pp. 38-56.
161. **Savourel, Yves** (2001) *XML Internationalization and Localization*. Indianapolis: SAMS Publishing.
162. **Schäler, Reinhard** (2002) "Can Translation Companies Survive the Current Economic Climate?" In proceedings of the International Conference *Translating and the Computer 24*, 21-22 November 2002, London. London: Aslib.
163. **Schäler, Reinhard** (1996) "Machine translation, translation memories and the phrasal lexicon: the localisation perspective". In proceedings of EAMT Workshop, Vienna, 29-30 August 1996.
164. **Seidel, John** (1998) "Qualitative Data Analysis". *The Ethnograph*, v5, Appendix E. Retrieved on 10/12/2007 from Qualis Research website: http://www.qualisresearch.com/.
165. **Sgarbas, Kyriakos, George Londos, Nikos Fakotakis and George Kokkinakis** (2003) "The WATCHER Project: Building An Agent for Automatic Extraction of Language Resources from the Internet". *Literary and Linguistic Computing*, Vol.18, No.4, pp. 449-464.
166. **Shuttleworth, Mark and Elina Lagoudaki** (2006) "Translation Memory systems: Technology in the service of the translation professional". In proceedings of *1st Athens International Conference of Translation and Interpretation*, 13-14 October 2006, Athens, Greece.
167. **Sidky, Ahmed S., Rajat R. Sud, Shishir Bhatia and James D. Arthur** (2002) "Problem Identification and Decomposition within the Requirements Generation Process." Technical Report TR-02-31, Computer Science, Virginia Tech.
168. **Simard, Michel** (2003) "Translation Spotting for Translation Memories". In proceedings of *NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*. Edmonton, Canada. May 31, 2003.
169. **Simard, Michel and Philippe Langlais** (2001) "Sub-sentential Exploitation of Translation Memories". In proceedings of *MT Summit VIII*, Santiago de Compostela, Spain.

170. **Somers, Harold and Gabriela Fernández Díaz** (2004) "Translation Memory vs. Example-based MT: What is the difference?" *International Journal of Translation* 16.2, pp. 5-33.
171. **Somers, Harold** (2003a) "Machine Translation: Latest Developments". In Ruslan Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press; pp. 512-528.
172. **Somers, Harold** (2003b) "Translation Memory Systems". In Harold Somers (ed.) *Computers and Translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing, pp. 31-47.
173. **Somers, Harold** (2001) "Bilingual Parallel Corpora and Language Engineering". Anglo-Indian workshop *Language Engineering for South-Asian languages (LESAL)*, April 2001, Mumbai, India.
174. **Somers, Harold** (1999) "Review Article: Example-based Machine Translation" *Machine Translation* 14, pp. 113-157.
175. **Sommerville, Ian** (2004) *Software Engineering*. 7th edition. Wokingham: Addison Wesley.
176. **Strauss, Anselm L. and Juliet M. Corbin** (1996) *Basics of qualitative research. Techniques and procedures for developing grounded theory*. Second Edition. London: SAGE Publications.
177. **Sutcliffe, Alistair** (2002) *User-centered Requirements Engineering: Theory and Practice*. London: Springer Verlag.
178. **Tardif, François and Anne Laugesen** (2006) "Going Global with TextBase TM Translation Memory". In proceedings of the International Conference *Translating and the Computer 28*, 16-17 November 2006, London. London: Aslib.
179. **Temponi, Cecilia, John Yen and W. Amos Tiao** (1999) "House of quality: A fuzzy logic-based requirements analysis". European Journal of Operational Research, issue 117.
180. **Thurmair, Gregor** (2003) "Making term extraction tools usable". In proceedings of the conference *Controlled language translation, EAMT-CLAW-03*, Dublin City University, 15-17 May 2003, pp.170-179.
181. **Townsend, Simon** (2007) "Over the waterfall". *ITNOW*, Magazine of the British Computer Society, January 2007.
182. **Ulrych, Margherita** (2002) "An Evidence-based Approach to Translation". In Alessandra Riccardi (ed.) *Translation Studies: Perspectives on an Emerging Discipline*. Cambridge: Cambridge University Press, pp. 198-213.
183. **Urquhart, Cathy** (2001) "An encounter with grounded theory: tackling the practical and philosophical issues". In Eileen Moore Trauth (ed.) *Qualitative research in IS: issues and trends*. Hershey/London: Idea Group Publishing.
184. **Valderrábanos, Antonio S.** (2004) "Who owns what? Some Insights on TM Ownership". On the website of TransRef - The Translation Reference Centre (URL: http://transref.org/default.asp?docsrc=/u-articles/TMownership1.asp - last accessed on 12/09/2008).
185. **Valderrábanos, Antonio S., José Esteban and Luis Iraola** (2003) "TransType2 – A new paradigm for translation automation". In the proceedings of *MT Summit IX*, New Orleans, Louisiana, USA, September 23-27, 2003.
186. **Varea, Ismael García, Daniel Ortiz, Francisco Nevado, Pedro A. Gómez and Francisco Casacuberta** (2005) "Automatic Segmentation of Bilingual Corpora: A Comparison of Different Techniques". In Pattern Recognition and Image Analysis. Berlin/Heidelberg: Springer.
187. **Vertan, Cristina and Vanessa Espin Martin** (2005) "Experiments with matching algorithms in example based machine translation". In proceedings of the International workshop: *Modern approaches in translation technologies, Borovets, Bulgaria, 24 September 2005*, pp. 42-45.
188. **Von Hippel, Eric** (1988) *The Sources of Innovation*. New York/Oxford: Oxford University Press.
189. **Wallin, Christina and Rikard Land** (2001) "Software Development Life Cycle Models - The Basic Types". Unpublished paper available at

http://www.idt.mdh.se/~rld/doc/Software%20Development%20Lifecycle%20Models%20The%20Basic%20Types.pdf - last accessed on 17/10/2005.

190. **Wallis, Julian** (2006) *Interactive Translation vs Pre-translation in the Context of Translation Memory Systems: Investigating the effects of translation method on productivity, quality and translator satisfaction.* MA dissertation. School of Translation and Interpretation, Faculty of Arts, University of Ottawa, Canada.
191. **Webb, Lynn** (1998) *Advantages and Disadvantages of Translation Memory: A Cost-benefit Analysis.* MA dissertation. Monterey Institute of International Studies, Monterey, Canada.
192. **Welsh, Elaine** (2002) "Dealing with Data: Using NVivo in the Qualitative Data Analysis Process". *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], *3*(2). Available at: http://www.qualitative-research.net/fqs-texte/2-02/2-02welsh-e.htm (last accessed on: 11/03/2008).
193. **Wheatley, Alan** (2003) "eCoLoRe (eContent Localization Resources for Translator Training). A Major Breakthrough for Translator Training". *The Globalization Insider*, Vol. XII, issue 2.4, 20 May 2003.
194. **Wilss, Wolfram** (1996) *Knowledge and Skills in Translator Behaviour.* Amsterdam and Philadelphia: John Benjamins.
195. **Xia, Yong** (2003) "Estimating Software Requirements". Seminar on Software Cost Estimation presented at Department of Cumputer Science, University of Zurich, Switzerland. Available at: http://www.ifi.unizh.ch/req/courses/seminar_ws02/reports/Seminar_8.pdf (last accessed on 12/09/2008).
196. **Yuste, Elia** (2004) "Corporate Language Resources in Multilingual Content Creation, Maintenance and Leverage". In proceedings of the 20th International Conference of Computational Linguistics COLING-2004, Geneva, Switzerland. Second International Workshop on Language Resources for Translation Work, Research & Training, pp. 9-15.
197. **Zanettin, Federico** (2002) "DIY Corpora: The WWW and the Translator". In Maia, Belinda / Haller, Jonathan / Urlrych, Margherita (eds.) *Training the Language Services Provider for the New Millennium*, Porto: Facultade de Letras, Universidade do Porto, pp. 239-248.
198. **Zerfaß, Angelika** (2002) "Evaluating Translation Memory Systems". In proceedings of LREC 2002: Language Resources in Translation Work and Research, 28 May 2002, Las Palmas de Gran Canaria.
199. **Zetzsche, Jost** (2006) *The Translator's Tool Box. A Computer Primer for Translators.* Oregon: International Writers' Group.
200. **Zielinski, Daniel and Yamile Ramírez Safar** (2005) "Research meets practice: t-survey 2005. An online survey on terminology extraction and terminology management." In proceedings of the International Conference *Translating and the Computer 27*, 24-25 November 2005, London. London: Aslib.

BIBL. LONDON UNIV.

Appendices

Appendix A: List of commercial Translation Memory systems

[correct as on 17 Sept. 2008]

1. ABREVE
Company: ArchiText
http://www.architext-usa.com/ABREVE.html

2. across
Company: Ahead Software AG
http://www.across.net/en/index.html

3. Anaphraseus
Developers: Open
http://sourceforge.net/projects/anaphraseus/

4. AppleTrans
Developer: Independent developers
http://appletrans.pbwiki.com/AppleTrans

5. Cafetran
Developer: Igor Kmitowski
http://www.cafetran.republika.pl/

6. Déjà Vu
Company: Atril
http://www.atril.com

7. Ecco
Company: PrimaTrans Software
http://primatrans.tripod.com/ecco/

8. ESTeam Translator
Company: ASTeam AB
http://www.esteam.se

9. Felix
Developer: Ginstrom IT Solutions
http://felix-cat.com/

10. ForeignDesk
Developer: Independent developers
http://sourceforge.net/projects/foreigndesk

11. Fusion Translate
Company: JiveFusion Technologies
http://www.jivefusiontech.com/products_FT.html

12. GlobalLink Translator Studio & TM Server
Company: Translations.com
http://www.translations.com/technology/technology-05.2.html

13. GlobalSight Ambassador
Company: GlobalSight Corporation
http://www.globalsight.com/index.html

14. Heartsome Translation Suite
Company: Heartsome
http://www.heartsome.net/EN/hsts.html

15. Lingotek
Company: Lingotek Inc.
http://www.lingotek.com

16. Logoport
Company: Lionbridge Technologies, Inc.
http://www.lionbridge.com/lionbridge/en-US/services/localization-translation/software.htm

17. LogoVista X PRO
Company: Logo Vista Corporation
http://www.logovista.co.jp/english/index.html

18. Masterin
Company: Master's Innovations Ltd.
http://www.masterin.com

19. MemoQ
Company: Kilgray
http://www.kilgray.com/kilgray/products/memoq

20. MemorySphere
Company: AppTek Inc.
http://www.apptek.com/products/index.html

21. MetaTexis
Company: MetaTexis Software and Services
http://www.metatexis.com/

22. MultiTrans
Company: MultiCorpora Inc.
http://www.multicorpora.ca/products/multiTrans4/

23. Omega-T
Developer: Omega-T development team
http://www.omegat.org/en/omegat.html

24. ProMemoria
Company: BridgeTerm
http://www.bridgeterm.com/en/promem.html

25. SDL TRADOS
Company: SDL International
http://www.sdl.com

26. SIMILIS
Company: Lingua et Machina
http://www.lingua-et-machina.com

27. Snowball
Company: Total Recall ApS
http://www.trmem.com/

28. Transit XV
Company: STAR AG
http://www.star-solutions.net/html/eng/produkte/TransitXV.html

29. TrAID
Company: Greek Institute for Language and Speech Processing
http://www.ilsp.gr/traid_eng.html

30. Transolution
Developer: Independent developers
http://transolution.python-hosting.com/

31. Wordfast
Company: Champollion WordFast Ltd.
http://www.wordfast.net

32. WordFisher
Developer: Környei Tibor
http://www.wordfisher.com/index.htm

Software localisation tools

1. Alchemy CATALYST 6.0
Developer: Alchemy Software Development Ltd.
http://www.alchemysoftware.ie/products/catalyst.html

2. AppleGlot
Developer: Independent developers
http://developer.apple.com/intl/localization/tools.html

3. CatsCradle 3.2
Developer: Stormdance
http://www.stormdance.net/software/catscradle/overview.htm

4. ENLASO Localization tools
Developer: Yves Savourel
http://www.translate.com/technology/tools/

5. Globalization Image Assistant (for localisation of images)
Developer: Bjorn Austraat of Austraat Globalization Solutions
http://www.imageglobalization.com/index.html

6. Globalyzer
Developer: Lingoport
http://www.lingoport.com/product/globalyzer

7. iLocalise
Developer: Arizona (Jean & Simon Bovet)
http://www.arizona-software.ch/applications/ilocalize/en/

8. LangBox (for localisation of C applications)
Developer: Gulf Computers, llc.
http://www.langbox.com/inter_e.html

9. Lingobit Localizer
Developer: Lingobit Technologies
http://www.lingobit.com/products/

10. Multilizer
Company: Multilizer Inc.
http://www.multilizer.com/

11. Open Language Tools (Sun)
Developer: Independent developers
https://open-language-tools.dev.java.net/

12. RC-WinTrans
Company: Shaudin
http://www.schaudin.com/products/products.htm

13. SDL Passolo
Company: SDL International
http://www.lspzone.com/en/products/software-localization/

14. Tramigo (for localisation of Macromedia Flash files)
Developer: Avral Technologies Limited
http://www.avral.com/tramigo/index.php?lng=en

15. TransLocale
Developer: Bantam Software
http://www.avral.com/tl/index.php?lng=en

16. Visual Localize
Developer: Applied Information Technologies AG
http://www.visual-localize.de

17. WebBudget XT
Developer: Aquino Software
http://www.webbudget.com/

18. WorldServer
Developer: Idiom Technologies Inc.
http://www.idiominc.com/products/

Appendix B: TM Survey 2006 web questionnaire (screenshots)

Translation Memories Survey 2006 Exit this survey

User profile (page 1/6)

Thank you for participating in the TM Survey 2006, conducted by Imperial College London.

The survey's goal is to strengthen the voice of translation professionals with regard to technology offerings for assisting the translation activity.

With your help, we hope, firstly, to find out to what extent and how Translation Memory tools are used today by translation professionals and, secondly, to discover ways of improving these tools so that they can be more practical, more user-friendly and, most importantly, meet better the needs of their users.

*** 1. What is your occupation?**

- ☐ Translator
- ☐ Subtitler
- ☐ Terminologist
- ☐ Project Manager
- ☐ Reviewer/QA Manager
- ☐ Other (please specify)

*** 2. How many years of experience do you have in this role?**

*** 3. What is your professional status?**

- ○ Freelancer working independently without an agency
- ○ Freelancer working closely with an agency
- ○ Freelancer working cooperatively with other freelancers
- ○ Company owner
- ○ Company employee (Translation/Localisation company)
- ○ Company employee (other sector, please specify)

*** 4. To which age group do you belong?**

- ○ <20
- ○ 20-30
- ○ 31-40
- ○ 41-50
- ○ >50

*** 5. Do you have a professional qualification?**

- ○ Yes
- ○ No

6. If yes, what type?

- ○ University degree
- ○ Certification by a private institution
- ○ Certification by a professionals' association
- ○ Other (please specify)

*** 7. How would you rate your general computer usage competence?**

- ○ Non-existent (No computer)
- ○ Poor
- ○ Adequate (Can send email, create WP docs)
- ○ Good
- ○ Excellent (Customise O/S, create macros)

Next page >>

Translation Memories Survey 2006 Exit this survey

Work environment & practices (page 2/6)

* 1. Where are you located?

* 2. What are the languages you work with?

* 3. Which operating system are you using?

- Mac OS
- Windows
- Linux
- Unix
- Other (please specify)

* 4. Do you have access to the Internet?

- Yes
- No

5. If yes, what kind of connection do you have to the Internet?

- 28.8 Kbps modem
- 56 Kbps modem
- ISDN
- Cable modem
- DSL
- T1 or better
- Do not know

* 6. What is the main reason you use the Internet for (in relation to your work)?

- e-mail (reception, processing and delivery of translation orders, communication with colleagues/clients)
- source of knowledge (information on specialised areas, reference texts)
- source of linguistic information (terminology, usage examples, contexts)
- download translation software
- download other software
- Other (please specify)

* 7. What is your main area of specialisation?

- Literature
- Legal translation
- Technical translation (incl. medical & scientific)
- Marketing material
- Other (please specify)

* 8. How many source words per year do you translate? If you are a freelancer, please give an estimate for your personal annual work (e.g. if you translate 3,000 words per day - working 5 days a week - your annual work will be around 780,000 words). If you are a company owner or employee, please give a rough estimate of the company's amount of translation work.

- < 1 million
- 1-5 million
- 5-10 million
- 10-20 million
- 20-50 million
- 50-100 million
- 100-500 million
- > 500 million

* 9. Which of the following format(s) do you mostly work with?

- Word/plain text/Excel/PowerPoint documents
- Hard copy documents
- HTML/XML webpages
- PDF documents
- FrameMaker files
- Other (please specify)

* 10. Does the content you usually deal with have...

	A lot	Average	Not that much	Not at all
internal repetition (same sentences are repeated in full or partially throughout a document)?				
external repetition (source text appearing in a new document has been encountered and translated before in another document)?				

11. If you have to translate a term or phrase and you don't know the translation (and doesn't exist in your TM database, if you use a TM tool), were do you usually look (in order of use)?

	First	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Last
in hardcopy dictionaries/glossaries											
in dictionaries/glossaries in CD-ROMs											
in old translations/glossaries that you have											
in printed texts on the subject											
on the Internet: online dictionaries/glossaries											
on the Internet: in a search engine (eg. Google)											
on the Internet: monolingual texts on the subject											
on the Internet: bilingual (parallel or aligned) texts on the subject											
on the Internet: ask questions in forums/communicate with other translators											
you ask a subject expert											
you ask a colleague											

<< Previous page Next page >>

Translation Memories Survey 2006 Exit this survey

Translation Memory systems: Usage (page 3/6)

* 1. Do you use any Translation Memory software package?

- Yes
- No

<< Previous page | Next page >>

Translation Memories Survey 2006 — Exit this survey

Translation Memory systems: Limitations of use (page 4/6)

*** 1. What are your reasons for not using a Translation Memory tool?**

- [ ] What's this? Never heard of it
- [ ] I've been told that they cost too much
- [ ] The one I would like to buy costs too much
- [ ] I've been told that it is not suitable for my work
- [ ] I have tried and evaluated one/several but I didn't like any
- [ ] Not suitable for my work
- [ ] I was using one in the past but it didn't bring any benefit to my work so I stopped using
- [ ] I have a TM software but I have not learned how to use it yet
- [ ] Other reason (please specify)

*** 2. Have you heard of any Translation Memory or Localisation tools? If yes, which one(s)?**

- [ ] I have not heard of such tools
- [ ] across
- [ ] Alchemy Catalyst
- [ ] AppleTrans (FREE)
- [ ] Cafetran
- [ ] CatsCradle (FREE)
- [ ] DejaVu
- [ ] Ecco
- [ ] ENLASO Localisation tools (FREE)
- [ ] ESTeam Translator
- [ ] ForeignDesk (FREE)
- [ ] Fusion Translate CMT
- [ ] Heartsome Translation Suite
- [ ] iLocalize
- [ ] Inlocate
- [ ] Lingobit Localizer
- [ ] LogiTrans
- [ ] Logoport
- [ ] LogoVista
- [ ] Masterin
- [ ] Memo Q
- [ ] MemorySphere
- [ ] MetaTexis
- [ ] Multilizer
- [ ] MultiTrans
- [ ] Omega-T (FREE)
- [ ] Passolo
- [ ] ProMemoria
- [ ] SDL TRADOS 2006
- [ ] SDLX
- [ ] SIMILIS
- [ ] STAR Transit
- [ ] Trados
- [ ] TrAID
- [ ] Translator's Intuition
- [ ] Transolution (FREE)
- [ ] Transware Ambassador
- [ ] Visual Localize
- [ ] WebBudget
- [ ] Wordfast
- [ ] Wordfisher (FREE)
- [ ] Other (please specify)

*** 3. Would you be willing to try out or do you plan to buy a Translation Memory software?**

- ○ No
- ○ Yes
- ○ Yes, I have one in mind

*** 4. If a TM software developer asked for your involvement when developing a Translation Memory software so that they build a system corresponding to your needs would you accept to participate in the process?**

- ○ Yes
- ○ No, because...

<< Previous page | Next page >>

Translation Memories Survey 2006 Exit this survey

Translation Memory systems: Usage and perceptions (page 4/6)

*** 1. What prompted you to use a Translation Memory tool?**

- Imposed by the company/translation agency I work for/am partner with
- My own choice
- My client's request

*** 2. How many years have you (if you are a freelancer) or your company (if you are a company owner or employee) used a Translation Memory software for?**

- less than a year
- 1-2 years
- 2-5 years
- 5-7 years
- 7-10 years
- over 10 years

*** 3. Which Translation Memory/Localisation tool(s) are you currently using?**

- across
- Alchemy Catalyst
- AppleTrans
- Cafetran
- CatsCradle
- DejaVu
- Ecco
- ENLASO Localisation tools
- ESTeam Translator
- Fusion Translate CMT
- Heartsome Translation Suite
- iLocalize
- Inlocate
- Lingobit Localizer
- LogiTrans
- Logoport
- LogoVista
- Masterin
- Memo Q
- MemorySphere
- MetaTexis
- Multilizer
- MultiTrans
- Omega-T
- Passolo
- ProMemoria
- SDL TRADOS 2006
- SDLX
- SIMILIS
- STAR Transit
- Trados
- TrAID
- Translator's Intuition
- Transware Ambassador
- Visual Localize
- WebBudget
- Wordfast
- Wordfisher
- My company's internal tool
- I have built my own
- Other (please specify)

*** 4. Is there any Translation Memory or Localisation tool that you are NOT currently using in you work, but you would like to?**

- Not any
- across
- Alchemy Catalyst
- AppleTrans
- Cafetran
- CatsCradle
- DejaVu
- Ecco
- ENLASO Localisation tools
- ESTeam Translator
- Fusion Translate CMT
- Heartsome Translation Suite
- iLocalize
- Inlocate
- Lingobit Localizer
- LogiTrans
- Logoport
- LogoVista
- Masterin
- Memo Q
- MemorySphere
- MetaTexis
- Multilizer
- MultiTrans
- Omega-T
- Passolo
- ProMemoria
- SDL TRADOS 2006

- [ ] SDLX
- [ ] SIMILIS
- [ ] STAR Transit
- [ ] Trados
- [ ] TrAID
- [ ] Translator's Intuition
- [ ] Transware Ambassador
- [ ] Visual Localize
- [ ] WebBudget
- [ ] Wordfast
- [ ] Wordfisher
- [ ] Other (please specify)

*** 5. Did you receive any training on using the Translation Memory software that you use?**

- ( ) No
- ( ) Yes, by my company/translation agency
- ( ) Yes, by the software developer
- ( ) Yes, attended a course/seminar
- ( ) Yes, the training was part of an academic course I did
- ( ) Other (please specify)

*** 6. What percentage of your translation work is processed using a Translation Memory tool?**

7. If it is anything less than 100%, please list the reasons why you do not use the TM tool for all your content?

*** 8. For which tasks do you usually use a Translation Memory tool?**

- [ ] Translation
- [ ] Quality Assurance: Translation Completeness
- [ ] Quality Assurance: Terminology checks
- [ ] Quality Assurance: Consistency checks
- [ ] Quality Assurance: checks for client's proprietary needs
- [ ] Analysis for quotation/invoicing
- [ ] Providing content to others within the organisation
- [ ] Terminology management
- [ ] Terminology extraction
- [ ] Comparing/merging projects
- [ ] Splitting TMs to create project memories
- [ ] Leveraging supplier TMs for in-house use
- [ ] Other (please specify)

*** 9. How do you fill in your translation memory database?**

- [ ] By aligning my legacy translations and their originals
- [ ] By downloading pairs of parallel texts from the Web and by aligning them
- [ ] By aligning a ready-made bilingual parallel corpus
- [ ] The database is empty and I fill it in as I translate
- [ ] By importing the client's TM
- [ ] By importing my agency's TM
- [ ] I use a remote TM database which exists on a remote server
- [ ] In another way (please specify)

*** 10. You voluntarily use a TM software because you believe that...**

- [ ] it improves the quality of translation
- [ ] it saves you time
- [ ] it reduces the cost of translation
- [ ] the consistency in terminology improves
- [ ] it is the best way to exchange resources (glossaries, TM databases)
- [ ] Other (please specify)

<< Previous page | Next page >>

Translation Memories Survey 2006 Exit this survey

Evaluation of Translation Memory systems based on users' experience (page 5/6)

*** 1. Which is the single Translation Memory/Localisation tool that you use the most?**

- across
- Alchemy Catalyst
- AppleTrans
- Cafetran
- CatsCradle
- DejaVu
- Ecco
- ENLASO Localisation tools
- ESTeam Translator
- Fusion Translate CMT
- Heartsome Translation Suite
- i.localize
- Inlocate
- Lingobit Localizer
- LogiTrans
- Logoport
- LogoVista
- Masterin
- Memo Q
- MemorySphere
- MetaTexis
- Multilizer
- MultiTrans
- Omega-T
- Passolo
- ProMemoria
- SDL TRADOS 2006
- SDLX
- SIMILIS
- STAR Transit
- Trados
- TrAID
- Translator's Intuition
- Transware Ambassador
- Visual Localize
- WebBudget
- Wordfast
- Wordfisher
- My company's internal one
- My own
- Other (please specify)

2. Which version (if you remember)?

*** 3. Please rate this tool in terms of...**

	Not satisfied at all	Poor	Quite satisfied	Very satisfied	Excellent!
(having enough features that you can use to do your work in a better way and the features are working as they are supposed to): Functionality					
(how often it crashes/ produces errors): Reliability					
(retrieving all available exact or fuzzy matches for a source segment): Efficiency in match recall					
(retrieving the correct exact or fuzzy matches for a source segment): Efficiency in match precision					
(searching & retrieving matches quickly): Efficiency in speed					
(how convenient and practical it is to use through its user interface): Usability					
(how easy it was for you to learn how to use it): Learnability					
(value for money): Price to usefulness (as a tool) ratio					
(support that you received from the developer when needed): Customer support					

4. Which feature do you love the most in the TM tool that you use?

5. Which is the most annoying thing that has happened (or happens a lot) to you, while using this TM tool?

<< Previous page | Next page >>

Translation Memories Survey 2006

Exit this survey

Future directions: Visualising the ideal Translation Memory tool (page 6/6)

*** 1. Most of the features below are currently being developed. Considering the translation(-related) tasks you perform every day, how important is to you for a Translation Memory tool to be able to...**

	Very important	Somewhat important	Not vey important	Not at all important	Don't know
handle PDF files					
support all languages (unicode)					
run on multiple platforms (MacOS, Linux, etc.)					
handle graphics (extract text from images/ flash movies)					
export its TM database in a standard format that will allow it to be imported and used efficiently by a different TM tool (comply with the TMX and SRX standards)					
plug in a Content Management System					
support project-specific translation/localisation directives					
have an integrated Optical Caracter Recognition (OCR) capability					
integrate spellcheckers for any language					
sort automatically lists in alphabetical order					
handle text in graphs and illustrations that are embedded in a text document					
search as you type (have results retrieved and displayed before finishing your typing)					
preview target full text before the translation is complete					
view full source text in original formatting while you translate					
display some context of the exact/ fuzzy match suggested					
construct and suggest fuzzy matches by combining different words/phrases (existing in your translation memory or in your terminology bank/lexicon), if an exact or fuzzy match cannot be found for a given source segment					

2. Which text processing environment do you find more practical to use while translating?

- Microsoft Word
- The TM's application environment
- Any, I use them interchangeably
- Other (please specify)

*** 3. Which software solution do you prefer?**

- One program with which you can perform all your translation-related tasks
- Several programs (offered altogether as a translation toolbox -or suite) each performing a different task (terminology management tool, concordance tool, etc.) that you can open and close depending on the task you have to complete

*** 4. Language Resources: How important is to you for a TM tool to be able to...**

	Very important	Somewhat important	Not vey important	Not at all important	Don't know
import dictionaries/ glossaries from a CD-ROM format					
locate bilingual parallel texts on your subject on the Web (you may specify authoritative sources only), download them, align them and fill in your TM database for future use or reference					
locate webpages on the Web that contain glossaries for your subject and language pair and bookmark them so that you can acess them through your TM tool					

5. Any comments for the above possible features? Is there any other particular task that you would like to see a Translation Memory do in the future?

<< Previous page | Next page >>

Translation Memories Survey 2006 Exit this survey

Thanks!

We appreciate your feedback. It is your ideas that will shape the future of Translation Memory systems.

Thanks again!

Elina Lagoudaki, PhD Researcher at Imperial College London

1. If you are interested in the findings of this survey, we can e-mail a copy of the results, when the survey is completed, to your e-mail address:

[ ]

<< Previous page | Done!

Appendix C: TM Survey 2006: Summary of quantitative results

Results Summary Show All Pages and Questions — Export... View Detail >>

Filter Results
To analyze a subset of your data, you can create one or more filters.
Add Filter... **Total:** 874 **Visible:** 874

Share Results
Your results can be shared with others, without giving access to your account.
Configure... **Status:** Enabled **Reports:** Summary and Detail

1. User profile (page 1/6)

1. What is your occupation?

	Response Percent	Response Total
Translator	**89.7%**	**784**
Subtitler	3.2%	28
Terminologist	3.8%	33
Project Manager	9.8%	86
Reviewer/QA Manager	6.6%	58
View Other (please specify)	8.9%	78
Total Respondents		**874**
(skipped this question)		0

2. How many years of experience do you have in this role?

View **Total Respondents**	**874**
(skipped this question)	0

3. What is your professional status?

	Response Percent	Response Total
Freelancer working independently without an agency	**47.9%**	**419**
Freelancer working closely with an agency	19.3%	169
Freelancer working cooperatively with other freelancers	5.7%	50
Company owner	7.9%	69
Company employee (Translation/Localisation company)	9.5%	83
View Company employee (other sector, please specify)	9.6%	84
Total Respondents		**874**
(skipped this question)		0

4. To which age group do you belong?

	Response Percent	Response Total
<20	0.3%	3
20-30	20.5%	179
31-40	**29.3%**	**256**
41-50	24.6%	215
>50	25.3%	221
Total Respondents		**874**
(skipped this question)		0

5. Do you have a professional qualification?

	Response Percent	Response Total
Yes	**89%**	**778**
No	11%	96
Total Respondents		**874**
(skipped this question)		0

6. If yes, what type?

	Response Percent	Response Total
University degree	**73.9%**	**584**
Certification by a private institution	3.8%	30
Certification by a professionals' association	14.3%	113
View Other (please specify)	8%	63
Total Respondents		**790**
(skipped this question)		84

7. How would you rate your general computer usage competence?			
		Response Percent	Response Total
Non-existent (No computer)		0.2%	2
Poor		0.1%	1
Adequate (Can send email, create WP docs)		5.4%	47
Good		**64%**	**559**
Excellent (Customise O/S, create macros)		30.3%	265
	Total Respondents		**874**
	(skipped this question)		0

2. Work environment & practices (page 2/6)

1. Where are you located?			
		Response Percent	Response Total
Albania		0.3%	2
Argentina		5.5%	43
Australia		1.4%	11
Austria		1.5%	12
Azerbaijan		0%	0
Belgium		1.2%	9
Bosnia and Herzegovina		0%	0
Brazil		4.9%	38
Bulgaria		0%	0
Canada		3.1%	24
Chile		0.8%	6
China		0.1%	1
Colombia		0.3%	2
Croatia		0.3%	2
Cuba		0%	0
Cyprus		0%	0
Czech Republic		0.4%	3
Denmark		0.4%	3
Egypt		0.4%	3
Estonia		0.8%	6
Finland		3.9%	30
France		4.4%	34
Germany		5.5%	43
Greece		10.1%	78
Hungary		0.9%	7
Iceland		0.1%	1
India		0.5%	4
Indonesia		0%	0
Iran		0%	0
Iraq		0%	0
Ireland		0.4%	3
Israel		0.3%	2
Italy		3%	23
Jamaica		0%	0
Japan		1%	8
Luxembourg		0.5%	4
Malaysia		0%	0
Malta		0.1%	1
Mexico		0.6%	5
Morocco		0%	0
Netherlands		2.1%	16
New Zealand		3.1%	24
Nigeria		0%	0
Norway		0.5%	4
Pakistan		0%	0
Poland		0.5%	4
Portugal		1.5%	12

Romania	0.3%	2
Russia	1.8%	14
South Africa	1.8%	14
Spain	3%	23
Sweden	3.7%	29
Switzerland	0.8%	6
Taiwan	0%	0
Turkey	0.6%	5
United Arab Emirates	0%	0
United Kingdom	**15.4%**	**119**
United States	9.5%	74
Uruguay	0%	0
Venezuela	0%	0
View Other country (please specify)	2.7%	21
Total Respondents		**775**
(skipped this question)		100

2. What are the languages you work with?

View **Total Respondents**	**774**
(skipped this question)	100

3. Which operating system are you using?

	Response Percent	Response Total
Mac OS	4.4%	34
Windows	**91.5%**	**709**
Linux	1.7%	13
UNIX	0%	0
View Other (please specify)	2.5%	19
Total Respondents		**775**
(skipped this question)		100

4. Do you have access to the Internet?

	Response Percent	Response Total
Yes	**99.7%**	**773**
No	0.3%	2
Total Respondents		**775**
(skipped this question)		100

5. If yes, what kind of connection do you have to the Internet?

	Response Percent	Response Total
28.8 Kbps modem	0.3%	2
56 Kbps modem	4.8%	37
ISDN	5.3%	41
Cable modem	18.6%	143
DSL	**58.2%**	**447**
T1 or better	7.6%	58
Do not know	5.2%	40
Total Respondents		**768**
(skipped this question)		107

6. What is the main reason you use the Internet for (in relation to your work)?

	Response Percent	Response Total
e-mail (reception, processing and delivery of translation orders, communication with colleagues/clients)	**87.3%**	**676**
source of knowledge (information on specialised areas, reference texts)	81.1%	628
source of linguistic information (terminology, usage examples, contexts)	76.5%	592
download translation software	29.3%	227
download other software	30.6%	237
View Other (please specify)	9.8%	76
Total Respondents		**774**
(skipped this question)		100

7. What is your main area of specialisation?			
		Response Percent	Response Total
Literature		3.5%	27
Legal translation		9.2%	71
Technical translation (incl. medical & scientific)		**60%**	**465**
Marketing material		8.6%	67
View Other (please specify)		18.7%	145
	Total Respondents		**775**
	(skipped this question)		100

8. How many source words per year do you translate? If you are a freelancer, please give an estimate for your personal annual work (e.g. if you translate 3,000 words per day - working 5 days a week - your annual work will be around 780,000 words). If you are a company owner or employee, please give a rough estimate of the company's amount of translation work.			
		Response Percent	Response Total
< 1 million		**72.7%**	**563**
1-5 million		17.3%	134
5-10 million		3.7%	29
10-20 million		1.6%	12
20-50 million		1.7%	13
50-100 million		0.6%	5
100-500 million		0.9%	7
> 500 million		1.4%	11
	Total Respondents		**774**
	(skipped this question)		101

9. Which of the following format(s) do you mostly work with?			
		Response Percent	Response Total
Word/plain text/Excel/PowerPoint documents		**95.7%**	**741**
Hard copy documents		18.9%	146
HTML/XML webpages		26%	201
PDF documents		43%	333
FrameMaker files		8.9%	69
View Other (please specify)		9.7%	75
	Total Respondents		**774**
	(skipped this question)		100

10. Does the content you usually deal with have...					
	A lot	Average	Not that much	Not at all	Response Total
internal repetion (same sentences are repeated -in full or partially- throughout a document)?	9% (71)	**45% (350)**	42% (321)	4% (31)	**773**
external repetition (source text appearing in a new document has been encountered and translated before in another document)?	12% (96)	**44% (340)**	38% (290)	6% (46)	**772**
				Total Respondents	**772**
				(skipped this question)	102

11. If you have to translate a term or phrase and you don't know the translation (and doesn't exist in your TM database, if you use a TM tool), were do you usually look (in order of use)?												
	First	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Last	Response Average
in hardcopy dictionaries/glossaries	20% (132)	**21% (139)**	11% (72)	10% (64)	12% (78)	7% (49)	6% (40)	3% (23)	4% (26)	3% (19)	3% (17)	**4.01**
in dictionaries/glossaries in CD-ROMs	**38% (228)**	16% (99)	8% (48)	6% (35)	6% (38)	5% (28)	3% (19)	3% (17)	3% (20)	6% (34)	6% (37)	**3.75**
in old translations/glossaries that you have	**17% (97)**	11% (65)	11% (64)	9% (54)	8% (44)	9% (53)	8% (46)	8% (47)	7% (39)	7% (41)	5% (31)	**5.12**
in printed texts on the subject	1% (4)	4% (18)	3% (16)	7% (32)	6% (27)	8% (39)	10% (47)	15% (67)	12% (57)	13% (60)	**20% (94)**	**7.80**
on the Internet: online dictionaries/glossaries	23% (159)	**23% (162)**	21% (147)	13% (93)	8% (55)	4% (25)	4% (25)	1% (10)	1% (8)	1% (6)	0% (2)	**3.08**
on the Internet: in a search engine (eg. Google)	16% (112)	**26% (189)**	25% (181)	19% (138)	6% (45)	4% (29)	1% (9)	1% (6)	0% (3)	0% (1)	0% (1)	**3.00**
on the Internet: monolingual texts on the subject	1% (7)	4% (22)	15% (95)	18% (112)	**20% (127)**	19% (118)	12% (77)	7% (42)	3% (16)	2% (10)	0% (2)	**5.20**
on the Internet: bilingual (parallel or aligned) texts on the subject	1% (5)	6% (35)	10% (62)	16% (96)	**21% (126)**	16% (95)	11% (65)	7% (40)	5% (27)	3% (20)	3% (20)	**5.55**
on the Internet: ask questions in forums/communicate with other translators	0% (2)	1% (4)	3% (18)	7% (46)	10% (62)	12% (77)	14% (89)	**15% (93)**	14% (87)	9% (56)	14% (84)	**7.53**
you ask a subject expert	1% (6)	1% (4)	2% (14)	4% (22)	6% (35)	10% (60)	11% (68)	17% (100)	17% (99)	**18% (106)**	13% (78)	**8.05**
you ask a colleague	1% (6)	1% (9)	3% (19)	5% (31)	7% (46)	9% (53)	16% (98)	15% (91)	**18% (110)**	15% (92)	10% (62)	**7.71**
											Total Respondents	**761**
											(skipped this question)	113

3. Translation Memory systems: Usage (page 3/6)

1. Do you use any Translation Memory software package?

		Response Percent	Response Total
Yes		**82.4%**	**635**
No		17.6%	136
	Total Respondents		**771**
	(skipped this question)		103

4. Translation Memory systems: Usage and perceptions (page 4/6)

1. What prompted you to use a Translation Memory tool?

		Response Percent	Response Total
Imposed by the company/translation agency I work for/am partner with		19.6%	123
My own choice		**72%**	**453**
My client's request		8.4%	53
	Total Respondents		**629**
	(skipped this question)		246

2. How many years have you (if you are a freelancer) or your company (if you are a company owner or employee) used a Translation Memory software for?

		Response Percent	Response Total
less than a year		10.5%	66
1-2 years		17.8%	112
2-5 years		**31.3%**	**197**
5-7 years		21.5%	135
7-10 years		13%	82
over 10 years		5.9%	37
	Total Respondents		**629**
	(skipped this question)		246

3. Which Translation Memory/Localisation tool(s) are you currently using?

		Response Percent	Response Total
across		2.4%	15
Alchemy Catalyst		7.8%	49
AppleTrans		1.1%	7
Cafetran		0.3%	2
CatsCradle		3.8%	24
DejaVu		22.9%	144
Ecco		0%	0
ENLASO Localisation tools		3.7%	23
ESTeam Translator		0%	0
Fusion Translate CMT		0.2%	1
Heartsome Translation Suite		1.6%	10
iLocalize		0.2%	1
Inlocate		0%	0
Lingobit Localizer		0%	0
LogiTrans		0.6%	4
Logoport		6.1%	38
LogoVista		0%	0
Masterin		0%	0
Memo Q		0.8%	5
MemorySphere		0%	0
MetaTexis		1.4%	9
Multilizer		1.6%	10
MultiTrans		1.8%	11
Omega-T		6.4%	40
Passolo		4.8%	30
ProMemoria		0%	0
SDL TRADOS 2006		24.4%	153
SDLX		19.4%	122
SIMILIS		0%	0

STAR Transit	13.5%	85
Trados	**50.6%**	**318**
TrAID	0.2%	1
Translator's Intuition	0.2%	1
Transware Ambassador	0%	0
Visual Localize	0.3%	2
WebBudget	0.3%	2
Wordfast	28.5%	179
Wordfisher	1.8%	11
My company's internal tool	2.9%	18
I have built my own	2.4%	15
View Other (please specify)	8.4%	53
Total Respondents		**628**
(skipped this question)		246

4. Is there any Translation Memory or Localisation tool that you are NOT currently using in you work, but you would like to?

	Response Percent	Response Total
Not any	**50.8%**	**319**
across	2.1%	13
Alchemy Catalyst	6.2%	39
AppleTrans	1.3%	8
Cafetran	0%	0
CatsCradle	0.5%	3
DejaVu	15.6%	98
Ecco	0%	0
ENLASO Localisation tools	1.1%	7
ESTeam Translator	0.2%	1
Fusion Translate CMT	0.3%	2
Heartsome Translation Suite	2.7%	17
iLocalize	0.3%	2
Inlocate	0%	0
Lingobit Localizer	0.2%	1
LogiTrans	0.5%	3
Logoport	1.4%	9
LogoVista	0.2%	1
Masterin	0%	0
Memo Q	0%	0
MemorySphere	0%	0
MetaTexis	0.8%	5
Multilizer	0.6%	4
MultiTrans	1.1%	7
Omega-T	2.4%	15
Passolo	4.8%	30
ProMemoria	0%	0
SDL TRADOS 2006	10%	63
SDLX	5.4%	34
SIMILIS	0.2%	1
STAR Transit	4%	25
Trados	4.8%	30
TrAID	0.2%	1
Translator's Intuition	0%	0
Transware Ambassador	0%	0
Visual Localize	0%	0
WebBudget	0.8%	5
Wordfast	5.3%	33
Wordfisher	0.5%	3
View Other (please specify)	4.5%	28
Total Respondents		**628**
(skipped this question)		246

5. Did you receive any training on using the Translation Memory software that you use?

	Response Percent	Response Total
No	**51.4%**	**323**
Yes, by my company/translation agency	11.9%	75
Yes, by the software developer	7%	44
Yes, attended a course/seminar	17.8%	112
Yes, the training was part of an academic course I did	7%	44
View Other (please specify)	4.9%	31
Total Respondents		**629**
(skipped this question)		246

6. What percentage of your translation work is processed using a Translation Memory tool?

View **Total Respondents**	**628**
(skipped this question)	246

7. If it is anything less than 100%, please list the reasons why you do not use the TM tool for all your content?

View **Total Respondents**	**456**
(skipped this question)	418

8. For which tasks do you usually use a Translation Memory tool?

	Response Percent	Response Total
Translation	**95.7%**	**601**
Quality Assurance: Translation Completeness	43.2%	271
Quality Assurance: Terminology checks	49%	308
Quality Assurance: Consistency checks	48.7%	306
Quality Assurance: checks for client's proprietary needs	19.3%	121
Analysis for quotation/invoicing	42.8%	269
Providing content to others within the organisation	7.8%	49
Terminology management	50.5%	317
Terminology extraction	23.9%	150
Comparing/merging projects	15.8%	99
Splitting TMs to create project memories	14.2%	89
Leveraging supplier TMs for in-house use	12.6%	79
View Other (please specify)	5.6%	35
Total Respondents		**628**
(skipped this question)		246

9. How do you fill in your translation memory database?

	Response Percent	Response Total
By aligning my legacy translations and their originals	51.3%	322
By downloading pairs of parallel texts from the Web and by aligning them	15.4%	97
By aligning a ready-made bilingual parallel corpus	12.4%	78
The database is empty and I fill it in as I translate	**73.4%**	**461**
By importing the client's TM	44.3%	278
By importing my agency's TM	28.2%	177
I use a remote TM database which exists on a remote server	6.5%	41
View In another way (please specify)	6.8%	43
Total Respondents		**628**
(skipped this question)		246

10. You voluntarily use a TM software because you believe that...

		Response Percent	Response Total
it improves the quality of translation		70.4%	442
it saves you time		**85.8%**	**539**
it reduces the cost of translation		34.2%	215
the consistency in terminology improves		82.8%	520
it is the best way to exchange resources (glossaries, TM databases)		31.4%	197
View Other (please specify)		10%	63
	Total Respondents		**628**
	(skipped this question)		246

5. Translation Memory systems: Limitations of use (page 4/6)

1. What are your reasons for not using a Translation Memory tool?

		Response Percent	Response Total
What's this? Never heard of it		10.4%	16
I've been told that they cost too much		10.4%	16
The one I would like to buy costs too much		13%	20
I've been told that it is not suitable for my work		1.9%	3
I have tried and evaluated one/several but I didn't like any		4.5%	7
Not suitable for my work		26%	40
I was using one in the past but it didn't bring any benefit to my work so I stopped using		3.2%	5
I have a TM software but I have not learned how to use it yet		14.9%	23
View **Other reason (please specify)**		**33.1%**	**51**
	Total Respondents		**154**
	(skipped this question)		720

2. Have you heard of any Translation Memory or Localisation tools? If yes, which one(s)?

	Response Percent	Response Total
I have not heard of such tools	11%	17
across	6.5%	10
Alchemy Catalyst	5.2%	8
AppleTrans (FREE)	3.2%	5
Cafetran	0.6%	1
CatsCradle (FREE)	8.4%	13
DejaVu	58.7%	91
Ecco	0%	0
ENLASO Localisation tools (FREE)	3.2%	5
ESTeam Translator	0%	0
ForeignDesk (FREE)	4.5%	7
Fusion Translate CMT	2.6%	4
Heartsome Translation Suite	3.2%	5
iLocalize	3.9%	6
Inlocate	0%	0
Lingobit Localizer	1.9%	3
LogiTrans	5.2%	8
Logoport	3.9%	6
LogoVista	1.9%	3
Masterin	0%	0
Memo Q	1.9%	3
MemorySphere	0%	0
MetaTexis	7.7%	12
Multilizer	1.9%	3
MultiTrans	16.8%	26
Omega-T (FREE)	9.7%	15
Passolo	10.3%	16
ProMemoria	0.6%	1
SDL TRADOS 2006	47.1%	73
SDLX	34.8%	54
SIMILIS	1.3%	2

STAR Transit		23.9%	37
Trados		**72.9%**	**113**
TrAID		1.3%	2
Translator's Intuition		1.3%	2
Transolution (FREE)		1.3%	2
Transware Ambassador		0.6%	1
Visual Localize		1.9%	3
WebBudget		1.9%	3
Wordfast		48.4%	75
Wordfisher (FREE)		7.7%	12
View Other (please specify)		7.7%	12
	Total Respondents		**155**
	(skipped this question)		719

3. Would you be willing to try out or do you plan to buy a Translation Memory software?

		Response Percent	Response Total
No		29.7%	46
Yes		**52.3%**	**81**
View Yes, I have one in mind:		18.1%	28
	Total Respondents		**155**
	(skipped this question)		719

4. If a TM software developer asked for your involvement when developing a Translation Memory software so that they build a system corresponding to your needs would you accept to participate in the process?

		Response Percent	Response Total
Yes		**69.7%**	**108**
View No, because...		30.3%	47
	Total Respondents		**155**
	(skipped this question)		719

6. Evaluation of Translation Memory systems based on users' experience (page 5/6)

1. Which is the single Translation Memory/Localisation tool that you use the most?

		Response Percent	Response Total
across		0.3%	2
Alchemy Catalyst		0.3%	2
AppleTrans		0.5%	3
Cafetran		0%	0
CatsCradle		0%	0
DejaVu		16.1%	95
Ecco		0%	0
ENLASO Localisation tools		0%	0
ESTeam Translator		0%	0
Fusion Translate CMT		0.2%	1
Heartsome Translation Suite		0.7%	4
iLocalize		0%	0
Inlocate		0%	0
Lingobit Localizer		0%	0
LogiTrans		0.2%	1
Logoport		0.2%	1
LogoVista		0%	0
Masterin		0%	0
Memo Q		0.7%	4
MemorySphere		0%	0
MetaTexis		0.8%	5
Multilizer		0%	0
MultiTrans		0.7%	4
Omega-T		2.5%	15
Passolo		0%	0
ProMemoria		0%	0
SDL TRADOS 2006		14.9%	88
SDLX		4.4%	26
SIMILIS		0%	0
STAR Transit		2.5%	15

Trados		**35.1%**	**207**
TrAID		0%	0
Translator's Intuition		0.2%	1
Transware Ambassador		0%	0
Visual Localize		0%	0
WebBudget		0%	0
Wordfast		17.5%	103
Wordfisher		0.3%	2
My company's internal one		0.2%	1
My own		0.7%	4
View Other (please specify)		1%	6
		Total Respondents	**590**
		(skipped this question)	285

2. Which version (if you remember)?

View **Total Respondents**	**496**
(skipped this question)	378

3. Please rate this tool in terms of...

	Not satisfied at all	Poor	Quite satisfied	Very satisfied	Excellent!	Response Average
(having enough features that you can use to do your work in a better way and the features are working as they are supposed to): Functionality	1% (8)	3% (18)	33% (192)	**47% (277)**	16% (94)	**3.73**
(how often it crashes/ produces errors): Reliability	2% (12)	7% (39)	**36% (213)**	35% (204)	21% (122)	**3.65**
(retrieving **all** available exact or fuzzy matches for a source segment): Efficiency in match recall	1% (5)	9% (55)	**46% (270)**	34% (198)	10% (61)	**3.43**
(retrieving the **correct** exact or fuzzy matches for a source segment): Efficiency in match precision	1% (5)	6% (37)	**45% (267)**	38% (224)	10% (56)	**3.49**
(searching & retrieving matches quickly): Efficiency in speed	1% (7)	6% (34)	26% (151)	**43% (254)**	24% (143)	**3.84**
(how convenient and practical it is to use through its user interface): Usability	1% (8)	7% (40)	29% (170)	**35% (209)**	28% (162)	**3.81**
(how easy it was for you to learn how to use it): Learnability	2% (11)	10% (58)	30% (179)	**36% (212)**	22% (129)	**3.66**
(value for money): Price to usefulness (as a tool) ratio	5% (27)	12% (70)	30% (174)	20% (115)	**34% (201)**	**3.67**
(support that you received from the developer when needed): Customer support	9% (52)	17% (98)	**39% (231)**	20% (117)	15% (91)	**3.16**
					Total Respondents	**589**
					(skipped this question)	285

4. Which feature do you love the most in the TM tool that you use?

View **Total Respondents**	**481**
(skipped this question)	393

5. Which is the most annoying thing that has happened (or happens a lot) to you, while using this TM tool?

View **Total Respondents**	**484**
(skipped this question)	390

7. Future directions: Visualising the ideal Translation Memory tool (page 6/6)

1. Most of the features below are currently being developed. Considering the translation(-related) tasks you perform every day, how important is to you for a Translation Memory tool to be able to...

	Very important	Somewhat important	Not vey important	Not at all important	Don't know	Response Average
handle PDF files	**55% (386)**	26% (182)	12% (87)	4% (27)	2% (17)	**1.72**
support all languages (unicode)	**40% (282)**	20% (137)	22% (154)	15% (107)	3% (20)	**2.21**
run on multiple platforms (MacOS, Linux, etc.)	19% (130)	17% (122)	**30% (207)**	29% (204)	5% (36)	**2.85**
handle graphics (extract text from images/ flash movies)	**37% (262)**	33% (233)	18% (127)	8% (55)	3% (23)	**2.06**
export its TM database in a standard format that will allow it to be imported and used efficiently by a different TM tool (comply with the TMX and SRX standards)	**53% (370)**	25% (175)	12% (85)	5% (34)	5% (36)	**1.84**
plug in a Content Management System	10% (73)	22% (155)	**27% (192)**	15% (102)	25% (177)	**3.22**
support project-specific translation/localisation directives	22% (155)	**33% (232)**	20% (139)	9% (65)	15% (108)	**2.63**
have an integrated Optical Caracter Recognition (OCR) capability	**28% (199)**	27% (188)	20% (140)	15% (103)	10% (69)	**2.51**
integrate spellcheckers for any language	**54% (374)**	24% (169)	12% (83)	8% (56)	2% (17)	**1.82**

sort automatically lists in alphabetical order	27% (188)	**29% (201)**	24% (166)	12% (86)	8% (59)	**2.47**
handle text in graphs and illustrations that are embedded in a text document	**57% (401)**	29% (205)	8% (53)	3% (21)	3% (20)	**1.65**
search as you type (have results retrieved and displayed before finishing your typing)	27% (191)	**34% (241)**	25% (177)	8% (58)	5% (33)	**2.29**
preview target full text before the translation is complete	**40% (280)**	35% (248)	16% (110)	4% (31)	4% (30)	**1.97**
view full source text in original formatting while you translate	**39% (274)**	36% (254)	16% (110)	6% (44)	3% (18)	**1.97**
display some context of the exact/ fuzzy match suggested	37% (256)	**42% (296)**	15% (102)	3% (20)	4% (26)	**1.95**
construct and suggest fuzzy matches by combining different words/phrases (existing in your translation memory or in your terminology bank/lexicon), if an exact or fuzzy match cannot be found for a given source segment	37% (257)	**41% (284)**	14% (97)	4% (31)	4% (31)	**1.99**
					Total Respondents	**699**
					(skipped this question)	175

2. Which text processing environment do you find more practical to use while translating?

		Response Percent	Response Total
Microsoft Word		**56%**	**388**
The TM's application environment		21.6%	150
Any. I use them interchangeably		19.6%	136
View Other (please specify)		2.7%	19
	Total Respondents		**693**
	(skipped this question)		182

3. Which software solution do you prefer?

		Response Percent	Response Total
One program with which you can perform all your translation-related tasks		**75.4%**	**528**
Several programs (offered altogether as a translation toolbox -or suite) each performing a different task (terminology management tool, concordance tool, etc.) that you can open and close depending on the task you have to complete		24.6%	172
	Total Respondents		**700**
	(skipped this question)		175

4. Language Resources: How important is to you for a TM tool to be able to...

	Very important	Somewhat important	Not vey important	Not at all important	Don't know	Response Average
import dictionaries/ glossaries from a CD-ROM format	**50% (351)**	30% (212)	13% (90)	5% (32)	2% (15)	**1.78**
locate bilingual parallel texts on your subject on the Web (you may specify authoritative sources only), download them, align them and fill in your TM database for future use or reference	**44% (305)**	31% (219)	16% (114)	7% (46)	2% (16)	**1.93**
locate webpages on the Web that contain glossaries for your subject and language pair and bookmark them so that you can acess them through your TM tool	**44% (307)**	34% (237)	15% (105)	5% (35)	2% (16)	**1.88**
					Total Respondents	**699**
					(skipped this question)	175

5. Any comments for the above possible features? Is there any other particular task that you would like to see a Translation Memory do in the future?

View	**Total Respondents**	**202**
	(skipped this question)	672

8. Thanks!

1. If you are interested in the findings of this survey, we can e-mail a copy of the results, when the survey is completed, to your e-mail address:

View	**Total Respondents**	**523**
	(skipped this question)	351

Appendix D: Excerpts from TM Survey 2006 responses on TM systems' strengths and weaknesses

RespondentID	Which is the most annoying thing that has happened (or happens a lot) to you, while using this TM tool?	Which is the single Translation Memory/Localisation tool that you use the most?	Which version (if you remember)?
250329381	often it crashes	DejaVu	
253151477	It does not update the TM automatically.	SDL TRADOS 2006	2005
253202001	Well, it crashes from time to time, but this is mostly because I was badly taught at the beginning. Now, I work with an instructor and if I meet a problem, he helps me. It is not the fault of the programme.	Trados	7
253207245	sometimes I know there is a match but it just doesn't recognize it.	Wordfast	5
253265665	it gets stuck very often	Trados	
253348194	when it crashes	Trados	
253383375	The format paint function is not always accurate and can be annoying.	SDLX	2005
253389159	limited amount of TM open at once	SDL TRADOS 2006	
253396211	losing the memory	Wordfast	
253434879	Clean-up didn't work and I was left with half a document to sort out...	Wordfast	5.25e
253547839	Importing errors	Trados	
253561056	Sometimes fuzzy matches are not applied correctly in my language, leaving certain terms untranslated.	SDLX	
253595657	formatting is not always retained, problems with tags	Trados	6.5
253607643	Differences in wordcounts.	Trados	6.5
253613695	sometimes I forget about the segment delimiters and that has caused me some trouble	Trados	
253614605	No direct support for Microsoft Word format; no internal spelling checker; no merge/split segment function	Omega-T	1.6.0 RC10
253622684	font and format changes, bad typographic quotes due to bad language attribution	Trados	6.5
253646106	some matches are not found because the Japanese handling is not 100% up to the task, but I really can live with that.	Omega-T	most of the time the latest beta. currently 1.6.0 RC12 cvs version
253648954	The format of the TM is not very well structured, so it is too easy for user errors to break the TM. Also, sometimes the software just does not work the way it should.	Wordfast	5
253651581	segmentaiton loss, TM mix up	Omega-T	always last beta when comming out
253659096	Degrading the editor's responsiveness when the material hits over a megabyte in size.	AppleTrans	1.1

253672852	It insists on using Japanese menus. The procedure for using a locale other than the system one is arcane and doesn't work 100%.	Omega-T	RC11
253697460	Nothing much... if I find something annoying, I log an RFE and if it is generally in demand, it gets implemented by the programmer.	Wordfast	5
253698742	Not wanting the export the file	DejaVu	no idea
253699691	Corrupt memories, changing font sizes, garbled formatting	SDL TRADOS 2006	7.5
253700459	It does not find a term in the termbase that I know is in there. Or finds a term that is similar but different.	Trados	6.5
253700991	It sometimes produces strange errors, or it crashes. It is also not 100% compatible with the TMX standard, which makes it difficult to import legacy TMs from third-party tools.	Wordfast	v5
253701205	Sth goes wrong with the programme and I cannot use it whereas the client expects me to	Trados	Only one I have
253701473	I hate it when it crashes for no reason	Wordfast	4.22t22
253701477	Change of tag format (which required revalidating TM entries)	Omega-T	1.6 RC10
253702033	Crashes	DejaVu	7
253702418	Rogue codes creating problems with export	DejaVu	DV3
253702493	DVX used to freeze on long segments - but this bug has been solved now	DejaVu	DVX 7.0.284
253703000	Problems with Greek fonts.	Trados	LSP 6.5
253704992	having to clean Word docs before importing (e.g. to get rid of large images) because otherwise it will slow the import process down and bloat the project file size	DejaVu	DVX 7.0.284
253705414	Occasional bugs, e.g. removing bullet points, section numbering, changing Styles.	SDL TRADOS 2006	7.5
253705603	Autopropagate doesn't always seem to work	DejaVu	DVX
253713721	Minimum score for database lookups not sufficiently adjustable or responsive Impossible to adjust minimum score separately for TM and terminology databases Segments with numbers assembled incorrectly (because of the way the terminology database treats numbers)	DejaVu	DVX v. 7.0.284
253716673	I cannot use it with large files (it starts taking seconds to a load, open and close a segemnt etc) - maybe I need more RAM (only using 256MB)	Transolution	0.4 beta 5
253716997	DVx is extremely poor at handling numbers	DejaVu	DVx. Pro 7.0.284
253717435	Truncated text after alignment	DejaVu	DVX
253717539	It can get a bit slow when retrieving stuff from the databases, but that may well be due to a) the size of my databases, which are now very large and b) an inadequately spec'd computer.	DejaVu	DejaVu 3
253718266	Segments in Legacy TMs are cut off (Excel length restriction)	DejaVu	
253724980	Term extraction tool from DejaVu can be better	DejaVu	DejaVu X
253725113	The need to use OpenOffice to convert files for use with OmegaT sometimes messes up formatting	Omega-T	1.6 RC10

253734590	some of the most simple editing functions don't work in a logical way (like there is no space between two words when you move a word to a differnet position in the sentence)	DejaVu	DVX
253740446	Codes spoiling matches	DejaVu	DVX
253742851	Terms not always sent to terminology database and/or project lexicon on command	DejaVu	DVX Build 284?
253746330	handles terms with apostropohes poorly	DejaVu	DVX 7.0.273
253747470	Lack of keyboard shortcuts	DejaVu	DVS 7.0.284
253750467	I have yet to figure out how to create new MultiTerm files. The program is so complicated that I haven't figured it out. I think the TRADOS customer support is absolutely below par given the price of the product.	Trados	7.1
253751272	processing speed at large source text files	SDLX	2005
253755922	database corrupted by other software	DejaVu	3
253757061	formatting problems	Trados	7.01
253757259	Every once in a while it 'hangs up' and does not retrieve matches although it knows they exist [source version displayed in the appropriate colour for, say, a 100% or 80% or whatever match]. This means the memory has to be 'reorganised', whereupon everything works again. This may however be a computer RAM related problem, I don't know.	Wordfast	4.22a
253759436	Mostly occurrences that were due to my insufficient Trados skills - at that time.	SDL TRADOS 2006	7.5
253761363	It often doesn't recognize matches because of font changes or internal formatting issues.	Trados	7.x
253767134	It crashes	SDL TRADOS 2006	6.5
253768493	Constant errors slowing me down Lack of error messages	SDL TRADOS 2006	2006
253788379	Search functionality only retrieves 99 hits, not more	Trados	5
253796306	when segments are assembled with embedded codes misplaced	DejaVu	DVX 7.0.284
253797155	Formatting in Word documents is corrupting	Trados	6.5.5
253797296	once in a while in large TMs it stops giving matches; I reported it (and others did) but it was never fixed, so it happens sometimes	Wordfast	5x
253803493	When using the Translate function in Trados with a file which has blocked (external) pretranslated segments, these pretranslated segments get read into memory as 'External' and all translations of these segments are lost. Since these are blocked (grayed out) segments in RTF/Word, Trados should be able to skip these rather than read them as 'External' segments where only codes show in Target.	Trados	7.1
253804481	Code problems in heavily (and/or inexpertly) formatted texts.	DejaVu	DVX
253815788	When codes (e.g. for formatting) are not correctly assigned in the target segment.	DejaVu	DVX build 284
253816423	Occasional bugs	Heartsome Translation Suite	6.1.4 (latest)
253819300	somewhat complex procedure for cleaning up dictionaries (related mostly to MS Access)	STAR Transit	Workstation XV SP 16

253821438	Wrong assembly!	DejaVu	X
253830491	While importing a new text the whole program crashed.	STAR Transit	XV update 17
253846140	As the PM, I have to manually enter all of the users and give them logins for each project. This gets really tedious with multiple 20 language projects.	Logoport	
253847065	rogue codes	DejaVu	DVX, most recent build
253871895	Inherent bugs due to legacy OmegaT code base. Some of these I have fixed and there are many more to fix in future.	Omega-t+	1.5M1 development version
253939769	Extremely sloppy tool, damages Word documents very easily causing formatting problems that call for repeated document re-creation. Lack of user warnings, error messages and tools compound the problem.	Trados	7.5.6
253954652	1. Doesn't find very similar text segments 2. Export doesn't work due to wrong codes	DejaVu	DVX
253996058	MS word's Normal.dot when changed by some other program or macro, MetaTexis corrupts and does not work, needs reinstalltion. Some features does not work for me.	MetaTexis	2.8
254015008	I have to check quite often the original in WYSIWYG and take care of italics and so on for which I cannot put the codes in the grid, Offending codes which should not really offend the system prevent exporting project and are time-consuming, DB maintenance takes too long as there are too many segments basically repeated to go through and just a code or mays differ...	DejaVu	DVX 7.0.284
254015495	problems with internal tags. poor flexibility with tags	DejaVu	273
254016645	some times it is unable to 'import' files which contain a lot of tables.	DejaVu	
254019856	Too many crashes when used with MS Word	Trados	6.5 Freelance Edition
254027738	Trados is what I call a 'clicking nightmare'. One has to click all the time when working in Trados, which is a real danger for our hands. I'm starting to suffer from RSI and I am convinced Trados has speeded up the process.	Trados	6.5
254032189	When working with Word files, formatting often gets corrupted, Greek fonts garbled, internal fields (for Table of Contents, etc.) corrupted. Also, I have not yet found a way to avoid unnecessary duplication of entries (first draft segment and corrected segment both remain in database sometimes). This may be due to my own incomplete knowledge of the software, but I consider Trados not easy to use to its full potential. Spell checking TagEditor files is extremely slow (Pentium 4, 1.80GHz, 512MB RAM, Windows XP)	Trados	Trados Freelance 6.0
254033307	It crashes all the time and retrieves, while translating, only a 60 percent of the 100% matches in the TM. It does not retrieve at all segments of one or two words.	Wordfast	5.25f
254037686	Too much use of memory, it slows down too much the PC	Trados	3
254041166	Formatting is changed	SDL TRADOS 2006	
254042673	translation folw attached to the segment - less freedom in langage expression -	Trados	6.5

254046225	Format changes	SDL TRADOS 2006	
254046572	it does not always show a match, and needs to be 'reminded'	Wordfast	3.5
254046615	To have to deal with the taglines in a sentence to be translated (i.e. for Framemaker)	Trados	6
254048929	messes up formatting	Trados	7
254055206	TM ceases to work for unknown reasons	Trados	5
254057576	Some TMs change the font and/or replace some signs such as accents per symbols. Sometimes it is difficult to work with large documents with many embedded pictures	Trados	5.5 and 6.5 (one in PC other in laptop)
254057940	the latest version is almost always bug-ridden, the programmer releases it exactly for the more daring users to experiment with it and give real-time feedback, in a day or less there will be a new release to fix that bug, etc. You need to be aware of what the latest stable release is. Another small thing that I don't know how to fix is that it returns 100% matches when the capitalization of the segment is different, so I have to block and capitalize or convert to small letters.	Wordfast	I use 5.16e. Never use the latest version, wait for the bugs to be removed
254058528	Problems with alignment	SDL TRADOS 2006	7.5
254060552	Wordfast is not 100 pct functional on Mac OS	Wordfast	5+
254060618	MS Word crashes	Wordfast	5
254062417	Some errors with fonts or any things which I am not sure how to fix and can't find the solution quickly.	Trados	7,0
254071018	Difficult to get the tags in Trados Tad Editor right!	Trados	7
254077380	The aligning takes too much time.	My own	2
254079770	Not finding a matching TU that is actually in the TM.	Wordfast	
254079876	When it does not clean the document properly	Wordfast	2003
254085752	freezes	Trados	6.5
254091749	The tags you cannot get rid of.	Trados	
254099649	failure to do text in textboxes	Wordfast	current new
254099994	System crash and data corruption.	Trados	3
254104009	losing format styles, working with tables in Trados may be a headache	Trados	6
254104382	Import resources do not work perfectly well	Trados	7
254113314	even though i know that i've translated something similar, i can't find using the tool.	SDL TRADOS 2006	2006
254126024	rare crashes with no warning	STAR Transit	XV service pack 17
254159648	Interface a bit complicated	DejaVu	DVX

254163472	The single most annoying thing about Trados is that things occasionally go wrong, Trados grinds to a halt, and it can be very hard to understand what the problem is and how to fix it. This has happened to me dozens of times. So far, I have always been able to recover, and whst's more without losing too much work, but it is a very hit and miss affair. The underlying problem to my mind is that the Trados documentation is not very good or comprehensive. That also applies to their support web site.	Trados	7
254163946	Inability to merge segments in most cases	Trados	6.5
254175774	Changes in layout when the source text is not produced usin Styles in -Word	SDL TRADOS 2006	
254179986	it sometimes cuts segments in an unusual way and makes a bit of a mess with my text.	Wordfast	
254180507	Hidden pictures not translated Text box hidden	Trados	6
254182145	Crashes, mysterious error messages	SDL TRADOS 2006	2006 Freelance
254187994	You have to use the mouse, because the cheap-ass GUI dialogs are not designed for keyboard-only navigation and operation. That's not only a pain in the ass, but downright absurd for a text-processing application, wouldn't you say? Secondary answers: It's a VBA app and can interfere with other Access/VBA applications, such as one timekeeping app I would have liked to use. It won't run on WINE (mainly a dongle issue, I gather).	DejaVu	7.0.284
254202383	The tool crashes as you finish the job or are cleaning the file and you loose everything. I have often had to clean a Trados document by hand.	MetaTexis	2.8? - auto updated regularly
254272656	Change of fonts in multilingual translations	Trados	6.5
254283276	Occasional freeze-ups	DejaVu	7.284
254284689	crashes	Alchemy Catalyst	6
254299600	Incompatibiity with an older system configuration	Wordfast	5.25c
254301044	DVX tends to crash more than DV3	DejaVu	DVX latest build
254317683	Poor design and functionality of TagEditor	Trados	6.5
254350967	global changes in a project containing multiple files/in a tm	SDL TRADOS 2006	2006
254355592	blocked glossaries	Wordfast	5
254358564	Lack of support for RTF	LocFactory Editor	1.6
254360800	Searching in the whole text after it has been segmented.	AppleTrans	1.1 (v32)
254377465	crashed tags	Trados	7
254382370	no spellchecking yet	Omega-T	almost always the last one (it changes fast)
254405105	Corrupts fonts regularly	Trados	7
254406802	It works better on PC (it's faster), but I prefer to work on my Mac.	Wordfast	5.22. Note: mostly on a PC, but also on my Mac.

254406824	bugs, not the most complete integration with Word	Wordfast	5
254409676	Lack of an auto-assemble tool to make it even faster.	AppleTrans	1.1 (32)
254409943	loss of Glossary importing	Wordfast	5.25
254415619	customer service responsiveness, no phone number available	SDL TRADOS 2006	
254472470	problems with number conversion	DejaVu	10
254591079	Incorrect segmentation, either because the program mistakenly breaks sentences up in the wrong place or because I accidentally mess up a segment markers. It's also tricky to work with bookmarks in Word documents.	Wordfast	Whichever is most recent
254591192	Some segments get pathological very occasionally and corrupt themselves when saved. Usually have to retype the sentence separately to get out of it. Might be Word's problem rather than Wordfast's, though.	Wordfast	3.35i
254615151	still too many bugs 2 yrs after launching	DejaVu	X 284
254625796	Delay in progression.	Omega-T	1.6.RC11b
254636054	No possibility to insert fuzzy/TM segment using a keybord shortcut.	Trados	6.5.5.439
254668336	The matching/assembly engine in DVX is still buggy, resulting in non-assembly, odd squares and crashes in some instances. GUI is excellent, but still has some rough edges. The company needs to improve on customer communication and a bit on support responsiveness.	DejaVu	3.x and DVX
254671348	Exporting the translation due to a problem with the Word file	DejaVu	7.0.284
254671971	Problems with Trados format files	DejaVu	x
254676214	poor error messages upon crashes or partial exports; ommission of an error message when unable to import a file and hogging larger and larger portions of virtual memory	DejaVu	X
254679622	Failure to translate 'non-editable' segments. Currently I have to carefully translated texts, particularly powerpoint presentations, to ensure that I have not missed non-editable text for translation.	Heartsome Translation Suite	6.1-4
254680265	corrupted tm	Trados	7
254682410	When clients urge me not to touch provided 100% matches	DejaVu	7.284
254682893	Small user interface problems (windows displaced) + incompatibility with other software (job management software, can't have both open simultaneously).	DejaVu	DejaVu X build 284
254689663	The Tag Editor is slow and not very user friendly.	Trados	6.5
254704270	Not being able to export the translation, corresponds to not being able to do a clean-up in Trados.	DejaVu	DVX 7.0.284
254714250	Terminology database user interface is bad.	DejaVu	DVX built 284 (latest)
254925426	Problems handling files with graphics	DejaVu	DVX 284

254954734	I sometimes get strange capitalization changes when inserting segments from the AutoSearch list. The most annoying thing though I suppose was breaking my dongle and having to buy another. However that expense is more than compensated for by the fact that Atril doesn't require us to purchase an upgrade every six months like Trados does.	DejaVu	DVX 7.0.284
254974962	The sw quitting on very long segments	DejaVu	DVX
254998200	Format paint feature	SDLX	2005 Standard
254999251	It sometimes refuses to open a new segment, tells that the previous segment isn't closed.	Trados	Freelance version 6
255019627	Export problems of finished translation - very occasionally	DejaVu	DVX
255026050	Some sentences disappeared, although I had translated them. It was a book and I didn't notice it, the editor saw it before printing it. Probably it happened because I was messing around with the tool, not being yet proficient.	Wordfast	5.22q
255216913	Failing to retrieve matches automatically which are there when I search for them	DejaVu	Pro
255560075	Matches often don't pop up right away. They have to be requested.	Wordfast	5.0zz2
255913030	Segments get 'blocked' and there is no way to move out of them, which renders the document unusable.	SDL TRADOS 2006	
256044514	Poor terminology management	Omega-T	1.6.0
256156997	TM provided by client is full of errors	Trados	6.5
256755918	There is no undo for extensive database operations	LocStudio	7.3
257033811	error that doesn't allow me to analyze a document or do the clean-up	SDL TRADOS 2006	7
257035458	TM functions slow down when strings have many data fields	Trados	7
257050929	tag damage	Trados	
257089776	Red colour for the terminology errors.	DejaVu	X
257120158	A long time ago, when I tried to use Trados,I could never get the screen where I wanted and finally one day, the whole machine crashed. I never tried to use it again. I hope to have better luck with WordFast	Wordfast	5.0a
257223143	crash, slow speed in finding matches	MetaTexis	latest
257253646	It may stop working at times, but not very often	STAR Transit	XV
257524983	some bugs	Trados	7
257682375	corrupt TTX files (sometimes) major formatting problems when using Workbench/Word	SDL TRADOS 2006	
258613923	Does not support RTF documents.	Omega-T	1.6
258638435	missing out matches due to different text format	Trados	6.5 lsp
258645771	FREQUENT CRASHES	Trados	TRADOS 6.5 LSP
258672148	The TM gets too big and you can't find a term that you know is there.	Trados	6.5

259028081	when the TM becomes too big, makes consulting too slow	Wordfisher	3.35X
259223735	Crashes	Trados	6.5
259232289	Crashes, doesn't clean up well all the time, throws out the formatting sometimes, adds extra 0s to numbered sequences... (gosh, why do I use it?)	Trados	6.5
259235006	too much on the screen; focus is on one sentence only - therefore no good in texts which are more demanding in terms of style	Trados	
259235347	When it won't process a file, and I have to go through hoops to make it work.	Trados	6.5
259235580	I keep finding the assembly of sentence portions a cumbersome process.	DejaVu	DVX
259235864	For some reason I've never been able to see the symbols at the top of the screen properly.	DejaVu	X
259237360	If I open another translation file at the same time, I can't access the same TMs that I am using on the first file (I have to close them first).	SDL TRADOS 2006	2006
259238640	Problems using Format Paint, when the structure of the English sentence is very different from the source text. They've given me workarounds, but I'd prefer a more elegant and reliable solution.	SDLX	2006
259242168	It crashes, though rarely.	Wordfast	
259246574	TagEditor crashing during Word spellcheck	Trados	7.1
259263059	Terminology management is dreadful with Multiterm.	Trados	7.1
259268748	It makes Word documents take much longer to save	Wordfast	
259273013	Not much. PDF documents remains a challenge, as well as powerpoints. But doing PPT's in Trados also takes a lot of work (not very quick.....)	Wordfast	
259293888	Problems with opening TM	SDLX	2006
259336064	Word used to crash quite often but it has not done so recently. I have backups and can usually sort problems when they occur.	Trados	6.5.5
259348129	Instability.	Wordfast	5.22
259356058	concordance does not work well in finnish due to case endings	Trados	7
259358006	Crashes. Requires 'Fix document' quite often because the TM says a TU is open, but I can't close it unless I exit from doc and memory and then fix doc.	Trados	7
259365587	In Trados, the impossibility of working easily with two target languages/dialects - e.g. US and UK English - and having both shown simultaneously in matches. When you have UK English as the target language, all the US English versions are blank in the Concordance window and vice versa.	SDL TRADOS 2006	7.5
259409985	Word interfaces alters formatting quite regularly	Trados	SDL TRADOS 2006
259417610	Crashing on returning to an already translated pair	DejaVu	7.0.284
259418681	When I accept a fuzzy match I lose my window layout and Enlgish takes over whole screen; I believe I am suposed t ocut & paste to not mess up formattign = laborious..	STAR Transit	XVII

259453791	unable to use Tag editor so far	Trados	Freelance edition 6.5
259496917	Problems with Powerpoint file - edits not transferred to ppt file - could just be me as a beginner though!	Wordfast	
259511792	It often cannot handle complicated layouts, and it keeps changing the font size. It throws up unintelligible error messages.	Trados	5.5
259548398	That when I finish one translation and start another I forget to change the project settings and the material in the TM is therefore stored with wrong file name, client etc. The reminder to set project settings only appears when I first open the TM and not when I open a new file to translate. (I only work with one TM for all my work.)	Trados	6.5
259575030	crashing; tm extraction not reliable, alignment feature prone to errors	across	
259581468	poor quality of service from Trados themselves, consistently unless you pay for it	Trados	
259586933	computer crash - does not happen very often, but more when using Wordfast than any other application I use on my Mac.	Wordfast	9
259638382	Loosing work because of the tool being unable to export the translated content!	SDL TRADOS 2006	7.5.0.756
259796742	there are many functions which I can't remember how to operate. so for instance I never add words to the glossary now as I'd have to search the instructions for how to do it. It's like a good programme where they gave up just as it was getting very good. When you start a translation for customer B you have to remember to change away from the TM for customer A	Wordfast	
259818097	Display does not keep up.	SDL TRADOS 2006	see above
259938500	Expanding segments often doesn't work - and when translating texts where segments have been expanded it often forgets about them. Also the way that is switches the source language settings in Word. I work in US English for one client and UK English for another - Trados seems to corrupt the settings in Word when it is triggered off so I constantly have to 'highlight all' then set the language - then remember to update the TM afterwards. The lack of flexibility in MultiTerm - it's fine for simple 1:1 meanings, but some customers use different terms and it would be useful to lump all the info into one screen. It's probably the way that I'm using it that's the problem :-)	Trados	6.5
259970317	apparent inconsistency in retrieving glossary terms	Wordfast	5
259973706	It sometimes crashes when a segment is open and causes problems. It is very easy to accidentally muck up the segment delimiters.	Wordfast	Sorry, I can't remember - it was the latest in November 2005
259997957	It crashes and it is slow	SDLX	
260327352	Missing paragraph markers.	SDL TRADOS 2006	Latest
260349009	Working with other translators: who manages the TM	Wordfast	
260351679	The continual offers of upgrading and the cost involoved	Trados	Freelance 7.5
260477159	difficulty with cleaning up the document; font and formatting of the target text change	Trados	6.5

260657293	clean up leaves unwanted junk in completed text	Trados	6.5
260752433	It's incomplete; export formats are not too impressive.	My own	
260983339	Tags inserted in the incorrect order in an otherwise good fuzzy match.	DejaVu	7.0.284
261207694	Crashing, time-wasting.	My own	
261407100	I got a new harddisk (and new OS) a few months ago. Since then, occasionally, for no reason I can establish, the clean up file appears empty instead of containing info. I haven't had the time or need to chase this since my clients usually ask for prior quotes, but if I were relying on the clean up file to know how many words, I'd translated, it would be a nuisance.	Wordfast	5.1
261410311	Forcing me to work on a sentence-by-sentence basis interrupts the flow of the language, and possibly degrades the quality of beauty of the language. Granted, this isn't a huge issue with highly technical texts, but once outside the realm of purely technical, style DOES matter. I also regreat that TM does not fit well into workflow when freelancers are editing/being edited. The final changes do not usually make it into the TM, since cleaned documents are sent for editing.	Wordfast	5.22
261410872	It crashes sometimes. Lack of support from vendor once you have owned the product for a while.	SDL TRADOS 2006	2006
261415053	I missed a deadline because of a bug in a filter.	Heartsome Translation Suite	6.2
261416004	macro crashes, Word hangs up	SDL TRADOS 2006	7.1
261416520	Using them at all.	Trados	
261416820	Nothing in particular.	DejaVu	
261422641	No option for using multiple TMs Difficulties in converting proprietory Trados file formats to other formats, (difficulties even in differenet Trados file versions)	SDL TRADOS 2006	FL 7.5.0.756
261425590	Sometimes there are bugs or hiccups that cause files not to be imported correctly into the tool.	STAR Transit	XV
261430036	the distinction between US and UK English occasionally means I can't use a client's TM - or have to transform it	SDLX	2005
261436472	Wrong translation coming up	Trados	7.0 Freelance
261442479	The conversion of pdf files that are not always recognized and must be manually imported from their converted .txt form, rather than being processed automatically like the .doc files. However, this happened in version 3.7. Version 4 was just released and it appears to have much better results, but I have only been using it 2 weeks now, which is not long enough to be certain.	MultiTrans	4
261460390	Workbench crashes due to 'illegal index', but I think I have to update some elements of my system in order to solve that. Will do that as soon as I have a little time...	SDL TRADOS 2006	2006
261467930	Cannot say.	SDL TRADOS 2006	
261468901	Difficult to find answer, help or solutions for some very simple practical questions. Manuals and help far too technical for a feelance translator who wants to keep it simple.	SDL TRADOS 2006	7
261470673	Fuzzy matches do not get displayed sometimes even though I am sure I have translated sth similar.	Wordfast	

261471057	When meaning is dependent on format, need to see original text.	DejaVu	3
261474351	Translations could not be cleaned or only after lengthy reformatting (Japanese Word styles seemed to collide with Trados).	Trados	5.5, but we'll upgrade to SDL Trados 2006 soon
261475871	Translation memory management should be better organized. TMs just grow and grow and the systems to keep them neat and tidy are not good.	Trados	we have several versions in use, mostly 5, 6 and 7.1
261476192	The TagEditor itself! It is neither a productive nor a practical tool at all!	SDL TRADOS 2006	7.5
261476367	spellcheck not working on target text after copying source into target box	Trados	5
261479932	segmentation rules not customizable as in previous versions. Was Ok in previous versions of SDLX.	SDLX	2005 SP1
261481620	The Logitrans component (of Logiterm) should be more flexible/sensitive. Fails to 'pick up' a lot of previously ued bitexts	LogiTrans	Version 3
261481778	That it is sometimes rather slow on our network	MultiTrans	4
261482128	mismatch	DejaVu	dvx7
261482553	Unappropriate/incorrect application of the Analysis and Translate features in Workbench; working with pre-translated segments slows down productivity (and there's no way to make client understand it!)	Trados	6.5
261482617	The way it treats tables	SDL TRADOS 2006	Freelance, SDL 2006
261483373	- When using Concordance search I often don't find the sentences I am looking for. I know that I have translated certain strings already but they don't come up. I usually have to change the words I am looking for. - Not being able to convert mif files st	SDL TRADOS 2006	SDL Trados 7.1
261483406	The bloody dongle messes me around (suddenly goes into demo mode or the dongle is not recognised on launching the program)	DejaVu	3
261483643	It slows down when the document or the TM are too big	SDLX	2004
261485843	The program crashed signficantly more when it was first released. It's greatly improved now, though, and is very stable in my opinion.	STAR Transit	Transit XV, service pack 18
261485916	crashes, totally unhelpful error messages, help system written for experts rather than 'mortal translators'	SDL TRADOS 2006	
261487955	(Non)Treatment of non-text segments Tag Editor	SDL TRADOS 2006	7.5
261488512	I find it difficult to use Multiterm (the term management database of SDL Trados). I often end up using Excel for glossaries.	SDL TRADOS 2006	SDL Trados 2006
261489348	crashes whilst analysing large Word files	Trados	7.1
261491464	No concordance match for capital letter acronyms	Trados	6.5

261491876	Cleaning up large Word documents with links, TOC, register, pictures, etc. can be a problem, even with a special macro supplied by Trados.	Trados	7.5
261496465	slows down on large documents, and messing up the curly brackets, I really hate that :-(	Trados	7.0 (bought SDL Trados, but I'm still not using it)
261496575	crashes with very big file sizes	Trados	6.5
261497816	Word integration	Trados	6.5
261503273	poor speed	Trados	5.5
261510097	Multiterm management is complicated. When transferring to another computer, it is hard to make the correct connections between termbase and Multiterm again (attach termbase).	Trados	5.5 freelance
261512936	It occasionnally gets its tabs wrong and adds bold to TagEditor files and such things	Trados	6.5
261518388	Cross-references in documents have altered the TUs and made them unusable.	SDL TRADOS 2006	SDL TRADOS 2006
261524430	Crashes of the TM on the server - deleting parts of the TM.	Wordfast	422 t22
261524456	Bugs that remain unfixed year after year, version after version.	Trados	6.5
261531268	Does not automatically open text contained in text fields in Word. Sometimes changes font type (e.g. Arial in source text to Times New Roman in translated text) Freelance version has relatively poor format conversion.	Trados	6.5
261536811	Deleting source segments in Word while working in Trados; also, annoying is that you can't work just in one user interface, but you have to work in Word AND Workbench or TagEditor AND Workbench etc.	SDL TRADOS 2006	
261539866	Terminology recognition is not consistent: punctuation as well as upper or lower cases can throw it off. An existing exact match is not found until I do a 'concordance search'.	Trados	v. 7.1
261545063	Produces errors very often.	Trados	6.5
261545944	Glitches with formatting.	SDL TRADOS 2006	7.5
261546264	In freezes the computer.	Trados	6.5
261547815	conflicts with some existing macros that disables the mouse.	Wordfast	5.25
261549842	incomplete error messages in the status bar (unable to read the whole thing) and being unable to troubleshoot it. Lack of support from supplier (without having to buy an expensive support contract). Poor knowledge database on the supplier's website.	SDL TRADOS 2006	7.x
261550720	no problems encountered that were directly caused by the software, but you sometimes have to be wary of having other software/templates active at the same time. Wordfast and Abbyy's software links in Word don't go together, but there's no real reason to need them to, so I run abbyy as a stand-alone programme.	Wordfast	5.25
261553881	Font issues for some languages when converting back to source format	SDLX	2006

261558563	Voice recognition software (Dragon) does not seem to like Wordfast and cursor can jump around the text	Wordfast	latest
261563470	Sometimes not finding matches I KNOW are in the TM.	DejaVu	DVX 7.0.284
261566505	With certain tagged formats, the result is sometimes strangely missing some tags. I therefore don't use SDLX for these any more.	SDLX	2004 Standard
261568131	It crashes often	DejaVu	DVX Pro
261571228	Sometimes the source segment cannot be seen without scrolling back in the document.	SDL TRADOS 2006	7.5
261573654	Deleted glossary entries for no reason in earlier versions	Wordfast	5.0h
261576023	It is doesn't work in all the languages. The company that built Translator's Intuition is coming out with a new product called Lingotek. I am eager to see how they improved it. I understand that it is web based with no software to install and they only charge a small monthly fee to use it. I can pay for it as I go along.	Translator's Intuition	
261580371	Having figures messed up e.g. percentages that are identical in my source and in a pair existing in the mdb, but the paragraph is numbered and DVX gets 'confused' and moves the paragraph number to the percentage, thereby messing it all up unnecessarily	DejaVu	7.0.273
261585371	I need to change windows constantly to access my electronic dictionaries.	SDL TRADOS 2006	2006
261585464	crashes, sudden corruptions of TM	SDL TRADOS 2006	7.5
261589585	Problems with pre-segmented texts, i.e. segments not being updated automatically etc.	Trados	7
261596391	System may lock up.	Fusion Translate CMT	
261620562	Slow at analyzing files Processes are all over the place I dislike the interfaces quite a lot	SDL TRADOS 2006	2006
261669116	Inability to open the tool at times!	Wordfast	
261670747	Non-matchng of glossary terms due to grammatical endings in the source language (German).	Wordfast	5.22b
261677308	font size/type changes in the target segment tool to translate Power Point files is really bad MultiTerm is too complicated (I never understood how the Project structure works) MultiTerm: it is sometimes not possible to change an input model MultiTerm: import is complicated and often fails for reasons that are not clear to me Text in an internal non-translateable style isn't always copied into the target text, but changed Now I notice all my frustration...	Trados	6
261688936	Although I have gotten over it, format painting is the most frustrating feature in SDLX	SDL TRADOS 2006	2006
261702752	Its interaction with Microsoft Word is very unstable.	Trados	6.5
261747837	it crashes often	Trados	7
261808019	when managing projects with traditional and simplified Chinese, for example when working with the Traditional text, TM match from the simplified TM is offered when no tradtional match is available, messing up the texts - there is a work-around though	DejaVu	DV X build 284

261815883	Cannot click twice fast on Translate/Get Next without crashing.	SDL TRADOS 2006	
261886885	- Crashes/Errors - Still not perfectly adapted to MS Word (e.g. language detection) - Contains no 'standard TM' covering most common terms (Yes/No, days, months, countries, figures etc.)	SDL TRADOS 2006	SDL Trados 7.5.0.756
261889230	information loss	Wordfast	3.35
261898180	bugs	Trados	7.1
261903525	program freezes	Trados	5.5
261906052	The 'translate-to-fuzzy-match' feature has a habit of halting on a match instead of riffling through the matches. There are a number of such 'issues', but I've learnt to live with them. I need the software to work with.	SDL TRADOS 2006	2006
261915947	Time required to read a PPT file in the Tag Editor and unexpected changes in Multiterm (language flags change without doing anything)	SDL TRADOS 2006	latest
261923296	broken tags that need fixing with the 'fix document' feature and changed fonts	Trados	6.5
261923462	Took a long time to master	Omega-T	
261923716	Can't match the sentences due to the [bad] formatting of the source file.	Trados	5.5
261949338	when Trados crashes	Trados	7.1
261952629	The re-integration of extracted terminology for translation (Send/receive function) does not work poperly.	STAR Transit	XV
261992467	OLE COM errors, layout and index problems after cleaning the files	Trados	6.5
262047792	I was using Wordfast with which I was quite familiar. I've started working with Trados/TagEditor just a couple months ago and I'm still learning to use it (regrettably, the hard way, i.e., on my own personal efforts), so I'm still not in a position to offer a positive contribution under this item.	Trados	7.1 Freelance
262226043	Trados often messes up formatting, so I have to go back and reformat after clean-up. That can be a real drag on large projects. Trados works very slowly on certain files that include numerous or large graphics or have special formatting (text boxes, frames).	Trados	5.1
262271816	Searching and replacing terms with Tag Editor, which is a tedious task.	Trados	6.5
262281437	That I removed all tranlsations one time because the translator did not know how to confirm/approve a sentence pair.	DejaVu	the latest version DVX
262405130	For long passages (> 15 words) LogiTerm can't distinguish between a close match and a perfect match.	LogiTerm	3.1
262549866	messed up cyrillic fonts, poor multiple translations merge, occasional errors when saving edited pieces into TM, copied text won't paste correctly into the target translation area (i.e. in MS Word) if extra space was selected. Have to go back and select the text from the first character of the sentence to the last character, and sometimes have to paste it as unformated text.	Trados	Trados 6,5 Freelance
262641397	aligning and term extraction can take quite a while	Wordfast	

262657883	I do not trust the TM tool blindly. So nothing annoying happens	Trados	6.5
262658106	When it maintains it's never seen a term before because it only appears near the end of a long segment.	DejaVu	DVX
262670575	Format painting	SDLX	Professional
262743131	Having to insert the embedded codes That some features don't always work properly That the system is incapable of recognizing the plural of a word that's been intered into the terminology database as a singular	DejaVu	7.0.384
262746672	1. Maintenance - > Filter When I combine three serch criteria and use both AND and OR it gets annoying. I cannot figure out the evaluation order. The help files does not give any guidance of which one takes priority? Example: Does 'xx AND yy OR zz' mean '(xx AND yy) OR zz' or does it mean 'xx AND (yy OR zz)' 2. It is better that the search string is case sensitive than if it was not. But the best would be if I could make the decision between case sensitive or not.	Trados	6.5
262866939	Crashes, especially with documents that contain odd formatting changes	SDL TRADOS 2006	the newest, 2006
262888017	sometimes it crashes MS word	SDL TRADOS 2006	7.5
263115416	Error: Can't process RTF file, though it was a doc.	SDL TRADOS 2006	7.5.0.756
263115667	crashing	Trados	6.5
263242570	the alignment feature is not as good as SDLX. The program crashes more than SDLX	DejaVu	DVX
263407201	A minor bug is introduced with a new feature. Of course, this goes with the territory if one wants a very user-driven up-to-date tool. Other tools may be more bug-free overall, but this also means their features (and bugs) are frozen in the software sometimes for several production cycles (can mean years).	Wordfast	Always the latest. Wordfast is updated on a constant basis.
263652880	computer resets and no changes saved neither in the TM nor in the file!	SDL TRADOS 2006	
263667620	crashes and inaccurate import of certain complex source documents	Memo Q	1
263704422	there is no undo function as yet, so it does happen that i press ctrl+z and loose what i translated in the segment.	Memo Q	
263922013	It doesn't work as advertised	Wordfast	constantly upgrade to the latest version (usually released every month or so)
263924516	problems learning how to deal with spaces at the end of sentences/segments. sometimes after export there are too few/too many spaces at the end of sentences.	DejaVu	DV10
263924603	Delimiters get deleted	Wordfast	5
263951457	User unfriendliness of file and information management, comments functionality etc	SDLX	
263957356	Bad segmentation, crashes.	Wordfast	5.25f

263984183	Dependency on other software running on PC - antivirus software, MS Office	Trados	Trados 6.5 Professional
264193945	Font corruption (diacrytics displayed wrong)	SDL TRADOS 2006	7.5
264221395	Sometimes I experience problems when cleaning up the target files.	Trados	5.5
264221975	It takes text out of context. Sometimes it is difficult to recognise and mantain format.	SDLX	2005
264226968	up grading	SDLX	Elite
264229398	V. few annoying features. One has to be careful to check for images which it skips over and which may contain material that needs to be translated.	Wordfast	5
264246517	I do not have the time to check out how to use it effectively and I presume that the reason it does not work very satisfactorily at times is that I am not using it properly.	Wordfast	
264255774	Some TU are recognized as 100% matches with the analysis application, but they appear as fuzzy matches after applying the TM.	SDLX	2006 (Build 6010)
264262072	before I really understand what means a MT and how usefull it can be, it was really annoying to try to perceive how this whole idea (segmentation, storage of translation solutions .. etc) can be usefull. Now I'm aware that the functionnality of TM's are not self-evident for neophytes ... they are not naturally attractive. And this is annoying: when something is good and we just don't catch how it can be usefull.	Wordfast	Wordfast 5.21i
264262451	Search and retrieve of matches is too slow.	Wordfast	5.1
264262995	Partial deletion of a glossary I had typed in by hand	Wordfast	5,35 (the latest)
264263049	PPT and xls files are difficult to process	Wordfast	Always the latest available
264263846	when it uses too much RAM	Wordfast	the latest
264270187	Conversion of fonts on Copy/Paste Corruption of Polish letters	Trados	6.5
264276544	I never have time enough to upgrade my knowledges.	Wordfast	5.1
264287809	Automatic spelling 'correction' when WW 'decides' the segment's language	SDL TRADOS 2006	Freelance
264326231	does not work in Excel, does not extract the exact matches	Trados	5.5, 6.5
264334537	sometimes the translation box is green, indicating there is an exact match in the TM but the box is empty and I have to use context search to retrieve the exact match text. another thing is that WF doesn't seem to recognize the difference between capitalized and small case - often the match is exact but in different case	Wordfast	5.16e
264336676	It crashes!	Trados	6.5
264338454	It sometimes does not retrieve a segment which I know perfectly well that I translated a day or two ago!	SDL TRADOS 2006	Freelance 2006
264351102	Influence use of MSWord editor in Outlook with regard to diacritic signs.	Wordfast	5

264352822	The fact that individual terms are sometimes not stored correctly or at all, especially if you are working in tables.	Wordfast	4
264377728	It cannot deal with tables with merged cells (only a slight problem, once you know it's likely to happen)	Wordfast	5.17y2
264388793	often crashes for no apparent reasons	Trados	7
264403569	Sometimes, I am sure of having translated a certain term/expression before, but I cannot retrieve it quickly for a new job.	Trados	6
264409231	Exporting bad/unreadable/incomplete Word documents	DejaVu	X
264414057	Sometimes terms that are indeed included in the TM are not retrieved when using the Concordance Search. For example, if you search 'Content' you don't get anything. But if you search 'Content Table' maybe you get something. Sometimes it seems to search by cluster, not by individual word.	Trados	5.5
264419503	After using the 'Clean up' function to clean translated files, the original format is very much altered and I have to devote a lot of time re-arranging the translated text, tables, figures, and so on.	Trados	7.0 Freelance
264424981	Creating the project itself. It is sort of a hassle because it involves many steps.	DejaVu	3.0.38
264431497	tags	Trados	7
264441074	Concordance does not work properly at times. Spell checker in TagEditor does not work at all.	Trados	6.5
264450248	Crashing Errors in retrieving terms for concordance	Trados	
264457376	Problems cleaning the document after translation. Rare, but annoying.	Trados	2007
264468203	I delete my segment delimiters by mistake, or they are overwritten and things go haywire.	Wordfisher	
264665587	an older version used to lose my id	Wordfast	5.15|1
264666269	Gives words that aren't supposed too be there.	Wordfast	
264666641	It sometimes refuses to recognise matches when segments are short	Wordfast	5.20n
264667148	Crashing	SDLX	SDL Edit 2006
264668880	Concordance search finds words according to match% which makes it sometimes hard to find terms in the TM if they are incorrectly spelled in the source text.	Trados	7
264671251	concordance does not find all relevant matches (e.g. because of different font etc.)	Trados	7.0 Freelance
264684870	Cannot think of anything really... There are some things that annoy me in TagEditor, I think it could be more user-friendly than it is at the moment, but otherwise I am very satisfied indeed.	Trados	7.1
264686496	The tool changes the text's format; the search options are quite basic - it would be very useful that Trados would let us search by target text instead of just source text. Some of the tools' applications could be much better and need development and bug fixes (e.g. TagEditor, T-windows, etc.).	Trados	6.5 Freelance
264699236	crashes, esp. caused by termstar.	STAR Transit	XV
264762421	Problems with segments (e.g. a segment cannot be closed)	Trados	6.5

264853716	Problems with the analysis or clean up of the dicument, or problems when trying to open a document with Tag Editor which, by the way, does not have a very friendly 'environment' (you cannot make any temporary mark in the document, or change the font colours and things like that.	Trados	6.5
264866255	Filters and other features are hard to master. Frequent crashed in older versions	Trados	7
264964419	Dividing segments in the wrong places.	Wordfast	5
265005380	I am flummoxed time and again by the following question stated in the negative form: Do you not wish to save the TM? Yes/ No	Wordfast	Beta version 07
265027173	Key combinations (macro) that don't work.	Wordfast	5
265073618	no match found for single-word segments	Wordfast	5.24a
265076018	It cannot handle certain fonts but replaces them with Times New Roman. In such texts, I have to use Wordfast because it doesn't screw up fonts and is still compatible with Trados, i.e. I can still use the uncleaned version and update my Trados TM with my Wordfast translations.	SDL TRADOS 2006	SDL Trados 2006
265118904	structured documents with footnotes do not work	Trados	6.6
265123046	It messes up the formulations of the original text, and attempts to translate tables often end with some catastrophe.	Trados	Freelance 6.5
265128053	Random errors and not being able to scroll down the hit screen without having the cursor there.	SDL TRADOS 2006	6.5
265144163	clients wanted Trados bilingual Word format	Memo Q	1.0.20
265234459	Lack of responsiveness with regards to development and upgrades	DejaVu	
265268212	TagEditor is very slow with texts of a certain client	Trados	SDL X 7.x
265576581	program errors, appearing in MS Word	Trados	Trados Freelance 7.0
265873462	conversion of a Trados memory into Wordfast	Wordfast	
265930314	No possibility of doing any kinds of global changes, all segments have to be modified one by one (e.g. replace full stops with commas etc.)	Clien'ts own tool	
265994777	Segmentation	Trados	7
266069547	I would like to choose how many words the segments I'm translating should contain.	SDL TRADOS 2006	Trados 7 Freelance
266214586	Having to select matching terms from a list	Memo Q	1
266507194	Alignment errors.	Trados	6.5
266761673	That it crashes, and I must start from scratch with a new file.	Trados	6.5
266828956	complains about open segments keep loosing font formatting (sometimes) more crashprone with certain OS:s and Word versions?	Wordfast	4,1
267265519	some files have bugs that suddenly cannot export a finished translation. I would like a function indicating that the file would not be exportable unless the bug was removed prior to import.	DejaVu	3.07
267278862	The versions when it makes Word crash.	Wordfast	5

267610732	Doesn't recognize fonts with special characters, sometimes too slow during analysis	Trados	6.5.2 or 7
267697840	Transit doesn't really have a concordance function. Trados has a strong tendency to change fonts. This is so annoying when your client has asked you to send uncleaned files.	STAR Transit	SP14
267721832	Since I work in Hebrew, occasionally there are files that become corrupt due to improper language settings in the source files.	SDL TRADOS 2006	7.5
268041523	Formatting problems in DTP programs, original formatting interfering with term recognition.	Trados	6.5 LSP
268236633	That it is not compatible with earlier versions.	Trados	7
268828512	It complains of errors every time I open it, about 5 times. It also has failed to import some Word documents, forcing me to use Trados instead.	DejaVu	X
268883856	Cannot process embedded objects (e.g. an Excel spreadsheet embedded in a PowerPoint slide)	SDL TRADOS 2006	7.5
269016576	When I know that i've already translated a phrase and it doesn't show	Wordfast	5.25f
269233848	No possibility to keep the files ordered in folders and subfolders when importing and exporting.	DejaVu	X
269561977	Concordance function is very annoying - does not produce latest hits - only old ones	Trados	6.5
270334164	Difficult to work long files....	Wordfast	Always the last one
271551415	Compatibility issue between Wordfast and DNS in recent versions of WF.	Wordfast	5
273088959	Change of typeface in External View documents	DejaVu	DVX
273521449	no tageditor	Wordfast	5.22
273554600	when it crashes	Trados	7
273568961	Application crash down without saving information	SDL TRADOS 2006	
273575995	Crushed	STAR Transit	XV
273584391	random tags	DejaVu	
273591316	WRONG EXPRESSIONS WERE USED FOR A TRANSLATION	SDL TRADOS 2006	DO NOT REMEMBER
273618024	Corrupted fonts (Greek characters)	Trados	Trados 6.5 Freelance
273642191	gibberish characters when using Greek	Trados	6.5
273674757	error messages, lost tags	Trados	7.0 Freelance
273696683	Font problems	Trados	6.5
273952390	it does not expand or reduce a segment	SDL TRADOS 2006	7
276518725	Works poorly on complicated MS Word files (e.g. destroys bookmarks and automated numbering, can't process 'heavy files' easily)	Trados	Trados 6.5 LSP / Freelance
280782135	Lack of speed	Trados	5.5
280998119	Change of layout	Trados	5.5

281076091	I had finish a translation and couldn't clean up the text!	Trados	6.5 Freelancer
281149213	Stored segments that are too long due to doubtful segmentation rules (e.g. a short word at the segment end which is interpreted as an abbreviation = several sentences together form a segment). Repeated subsegments in translatable text aren't found automatically although they are in the TM for sure.	Trados	7
281163101	Crashes	Trados	6.5
281410975	Efficiency in speed	Wordfast	
281468089	When I upgrade to a higher version, something always stops working on my computer.	SDL TRADOS 2006	7.5
281489191	Mostly codification errors	SDL TRADOS 2006	7
282279278	The spellchecker in Tag Editor is extremely slow and crashes all the time.	Trados	6.5 Freelance
284524907	It somtimes conflicts with the operating system.	Wordfast	5
284747232	When trying to locate errors in tags in rtf files with the use of S-tagger for FrameMaker. Spellchecking a ttx file.	Trados	6.5 FL
284783367	When importing txts, the target language characters are distorted	Trados	7
285150990	When working with different clients' 'platforms', things get mixed up, and I do most al of it by feding data one-by-one.	Trados	Trados 97
285839466	database corrupted	Trados	7
286938569	non-leggibility of Cyrillic fonts	SDL TRADOS 2006	6.5

RespondentID	Which feature do you love the most in the TM tool that you use?	Which is the single Translation Memory/Localisation tool that you use the most?	Which version (if you remember)?
250329381	the lexicon and the auto assemble	DejaVu	
253151477	Propagation of terms	SDL TRADOS 2006	2005
253202001	That it forces me to work steadily and consistenly through a text. Before, I used to 'jump' difficulties and leave them to solve later on. With Trados I have to work each one as it shows up - this is very good. Also, terminology remains consistent, and	Trados	7
253366167	ease of use - works in word	MetaTexis	
253366492	Concordance	Trados	
253383375	The separation of ST and TT into two columns	SDLX	2005
253389159	sentence level segmentation reliability speed correct export of translated text to the file format	SDL TRADOS 2006	
253396211	can be used without knowing too much about it	Wordfast	
253547839	Allignement	Trados	
253561056	Interface	SDLX	
253595657	easy to use	Trados	6.5
253613695	100% matches! and the possibility to go paragraph by paragraph seeing the source and target segments	Trados	
253614605	The (term or phrase) Find function	Omega-T	1.6.0 RC10
253622684	good ergonomy, speed	Trados	6.5
253646106	ease of use	Omega-T	most of the time the latest beta. currently 1.6.0 RC12 cvs version
253648954	It helps us get work done quickly while ensuring consistency among a wide variety of tasks.	Wordfast	5
253651581	no price, very good functionality, good helpline	Omega-T	always last beta when comming out
253659096	Flexibility in defining segmentation rules.	AppleTrans	1.1
253672852	Price	Omega-T	RC11
253691910	Flexibility for treating various file formats.	DejaVu	DV3
253695125	Autoassemble and terminology use, desinged for transalators	DejaVu	DVX always newest Beta

			version
253696126	That I can handle more than one text in one project	DejaVu	DVX
253697460	Ease of use	Wordfast	5
253698742	Simple to use	DejaVu	no idea
253699691	The concordance function	SDL TRADOS 2006	7.5
253700459	Concordance	Trados	6.5
253700991	Concordance and context search	Wordfast	v5
253701477	Simplicity	Omega-T	1.6 RC10
253702033	Automatic assembly of TM matches and terminology	DejaVu	7
253702418	Autoassemble	DejaVu	DV3
253702493	1. Assemble 2. Row Selector 3. SQL in project/TDB/MDB 4. Lexicon	DejaVu	DVX 7.0.284
253703000	The fact that is used within the MSWord application.	Trados	LSP 6.5
253704992	hard to decide - probably AutoAssemble (assembling of sub-segment portions found in the databases as a basis for entering the translatoin), or scanning the memory database (concordance search), or filtering on an expression to have only those segments in	DejaVu	DVX 7.0.284
253705414	General quality of translation work	SDL TRADOS 2006	7.5
253705603	Clear, uncluttered screen layout	DejaVu	DVX
253713721	Auto-assemble segments from terminology database as well as TM	DejaVu	DVX v. 7.0.284
253716673	Stability, clarity of display, ease of work with display	Transolution	0.4 beta 5
253716997	Autoassemble	DejaVu	DVx. Pro 7.0.284
253717435	Assemble and EBMT functions	DejaVu	DVX
253717539	The sub-sentence Assemble feature (ability to put sentences together from bits and pieces elsewhere in the databases, and even to turn fuzzy matches into exact matches if the correct translations are in the terminology base). I don't get much repetition	DejaVu	DejaVu 3
253718266	Autoassemble, Autopropagate	DejaVu	
253724980	External view for spellchecking	DejaVu	DejaVu X
253725113	Its speed and simplicity	Omega-T	1.6 RC10
253734590	substitution of terminology in the target segment on the fly	DejaVu	DVX
253740446	The structure and control it provides over language resources	DejaVu	DVX
253742549	Pretranslate	DejaVu	Professional

253742851	Assembly - assemble target text from smaller matching pieces of source text from TM. Thus possible to have a 'match' even if the exact source text is not in the TM.	DejaVu	DVX Build 284?
253746185	Ability to translate many files at the same time with different formats, e.g. word, excel, powerpoint files	DejaVu	Deja Vu X Workgroup
253746330	can add (and occasionally delete) terms as I work	DejaVu	DVX 7.0.273
253747470	Auto Search and Auto Assemble	DejaVu	DVS 7.0.284
253750467	Concordance search	Trados	7.1
253751272	Alignment	SDLX	2005
253755922	Speed and reliability	DejaVu	3
253757061	concordance search	Trados	7.01
253757259	Value for money esp. in terms of ease of learning.	Wordfast	4.22a
253759436	Concordance	SDL TRADOS 2006	7.5
253767134	Concordance	SDL TRADOS 2006	6.5
253782436	The alignment tool. I really missed it when I had to use Star Transit for a project recently.	Trados	Version 6.5 Freelance
253788379	Copy source segment	Trados	5
253796306	Filtering on selection	DejaVu	DVX 7.0.284
253797155	Time saving	Trados	6.5.5
253797296	term blacklist - improves consistency	Wordfast	5x
253802113	usability and learnability	DejaVu	284
253804481	1) Being able to filter a project (even consisting of multiple files in multiple formats) for a term or expression so that all instances of the term/expression in the project are displayed at once, which allows me to choose the best option for said term t	DejaVu	DVX
253815788	Assemble from portions (i.e. matches at the sub-sentence level)	DejaVu	DVX build 284
253816423	The fact that it runs on Macintosh	Heartsome Translation Suite	6.1.4 (latest)
253819300	Concordance search	STAR Transit	Workstation XV SP 16
253821438	Assembly	DejaVu	X
253827628	Ha! The money it earns me, I guess.	Trados	7.0 (Trados is terribly programmed software; I hate upgrading.)
253830491	It is so fast!	STAR Transit	XV update 17
253847065	Auto-assemble	DejaVu	DVX, most recent build
253871523	Assemble function	DejaVu	DVX 284

253871895	Docking user interface that allows user to configure their set up the way they like. Configurable fonts to make use easier on the eyes. These features are not currently available in OmegaT.	Omega-T+	1.5M1 development version
253882233	its terminology recognition capabilities	Wordfast	5.25
253936538	Align	Trados	7
253939769	Concordance	Trados	7.5.6
253954652	Terminology look up / matching	DejaVu	DVX
253996058	It works inside MS word as macros. My company's almost all Docs are perpared in this. So output translated doc's layout remain same as it was it in original. No need to redesign/re-layout the output.	MetaTexis	2.8
254015008	Scan-Compare, Propagate, Change Order of translation units	DejaVu	DVX 7.0.284
254015495	user interface	DejaVu	273
254019856	Word Count feature	Trados	6.5 Freelance Edition
254032189	The possibility to retrieve previous translations of terms within context, saving time and ensuring consistency within job and within bulk of work for particular client.	Trados	Trados Freelance 6.0
254033307	The Search feature, the various shortcuts to do things quickly and the user friendly interface. This is a very quality oriented tool.	Wordfast	5.25f
254041166	Concordance	SDL TRADOS 2006	
254041562	It's simplicity and it's accuracy.	Wordfast	5.18g
254042673	legacy	Trados	6.5
254046225	F3: find individual words	SDL TRADOS 2006	
254046572	Its simplicity and lack of 'Bloat'	Wordfast	3.5
254046615	Ability to translate Excel files quite easily	Trados	6
254048929	consistency	Trados	7
254057576	Retrieving matches and retrieving your work if the Word file crashed.	Trados	5.5 and 6.5 (one in PC other in laptop)
254057940	Ease of use closely followed by support closely followed by price	Wordfast	I use 5.16e. Never use the latest version, wait for the bugs to be removed
254058528	Variety of formats that can be translated	SDL TRADOS 2006	7.5
254060174	Time saver	Wordfast	
254060552	it does its job, mostly	Wordfast	5+
254060618	mouse-free usage and simplicity.	Wordfast	5

254077380	I built it myself (around 1990) and it cost me very little.	My own	2
254079770	I love its price and its ease of learning.	Wordfast	
254079876	Retrieval of phrases already translated	Wordfast	2003
254085752	concordance tool	Trados	6.5
254091749	Makes the translation process faster	Trados	
254092628	Reliability, usability, price, customer support and not being a bulky program. Would be even better if compatible with Linux and OpenOffice.	Wordfast	3.35
254099649	analysis of new text against an existing TM and the yahoo forum - v friendly	Wordfast	current new
254099994	It does what TM software is supposed to do.	Trados	3
254104009	concordance	Trados	6
254104382	MultiTerm	Trados	7
254113314	automatic input	SDL TRADOS 2006	2006
254126024	ease of use in not having to work in MS Word	STAR Transit	XV service pack 17
254159648	Import/export of data	DejaVu	DVX
254163472	As mentioned in Q 10 on the previous page, my favourite feature is simply the physical presentation of segments (in Trados, in two differently coloured boxes one above the other, with differently coloured backgrounds). The usefulness of other facilities (	Trados	7
254163946	Concordance search	Trados	6.5
254175774	Concordance	SDL TRADOS 2006	
254180507	Search feature	Trados	6
254187994	Cutting and joining segments as desired! Yeah! Must-have!	DejaVu	7.0.284
254202383	Easy to use. small learning curve, excellent support. (I used Trados almost exclusively for years and it took many months, with almost no support, to learn the rudimentaries. After 3-4 years I was using it to 10% of its potential. Not the case with MetaTe	MetaTexis	2.8? - auto updated regularly
254283276	Auto-Assemble	DejaVu	7.284
254284689	file management for version control	Alchemy Catalyst	6
254299600	Learnability	Wordfast	5.25c
254301044	easy to use, great customer support, great for multiple file projects, a must for XML and Powerpoint	DejaVu	DVX latest build
254317683	Concordance	Trados	6.5
254355592	usability	Wordfast	5
254358564	Being able to write scripts for it.	LocFactory Editor	1.6

254360800	The possibility to cross more than two languages at a time. Totally Unicode friendly The possibility to add color, italic, bold, etc. The really clear interface.	AppleTrans	1.1 (v32)
254369594	Affordability	Wordfast	5.1
254377465	Pre-translation	Trados	7
254382370	simplicity	Omega-T	almost always the last one (it changes fast)
254405105	Segmentation	Trados	7
254406802	It's easy to make my own glossaries.	Wordfast	5.22. Note: mostly on a PC, but also on my Mac.
254406824	segmentation in word processor - excellent respect of original format	Wordfast	5
254409676	The speed	AppleTrans	1.1 (32)
254409943	Easiness	Wordfast	5.25
254415619	concordance	SDL TRADOS 2006	
254472470	fuzzy matches	DejaVu	10
254591079	I love that it works on a Mac. I also love the price/performance ratio and the user support.	Wordfast	Whichever is most recent
254591192	It makes working in Word bearable. (Hate Microsloth's cruddy software, the last decent word processor they made was 5.1/mac.)	Wordfast	3.35i
254615151	Autoassemble (user-controlled machine translation)	DejaVu	X 284
254625796	Layout.	Omega-T	1.6.RC11b
254636054	Built-in glossary.	Trados	6.5.5.439
254668336	translation assembly matching according to subject hierarchy and client attributes, and the guaranteed matches in the DVX Workgroup edition	DejaVu	3.x and DVX
254671348	Pretranslating	DejaVu	7.0.284
254671971	AutoAssemble All format in one interface	DejaVu	x
254676214	single interface for different file formats in the same project; client specific TM's when I want so; usage of more than one TM; auto-assemble	DejaVu	X
254679622	Multi-platform nature and complete adherence to open standards	Heartsome Translation Suite	6.1-4
254682410	Ability to store and correct terminology on-the-fly	DejaVu	7.284
254682893	Assemble function, i.e. feature that is capable of retrieving portions from the MT to give a match in a new TU.	DejaVu	DejaVu X build 284
254689663	Concordance	Trados	6.5
254704270	Auto-Assemble, Filter function	DejaVu	DVX 7.0.284

254714250	Ability to load more than one file as one project.	DejaVu	DVX built 284 (latest)
254868271	AutoAssemble (assembling the translation proposal on the basis of glossary and fragments of TM segments)	DejaVu	284
254891740	Multitrans' unique corpus feature, different from a tradition 'translation memory'	MultiTrans	4
254925426	Ease of use	DejaVu	DVX 284
254954734	AutoSearch. Even if I don't use DVX's auto assemble feature in a segment, I love having my saved terms right in front of me. I also love that I can import different file formats into the same project and view those segments together all in one place.	DejaVu	DVX 7.0.284
254974962	The 'assemble' function	DejaVu	DVX
254999251	The concodance	Trados	Freelance version 6
255019627	Scanning the TM while translating to retrieve exact matches, fuzzy matches and ideas for solutions	DejaVu	DVX
255026050	I can verify the consistency of the translation very easily	Wordfast	5.22q
255065232	Simplicity (but virtually all suffer from featuritis)	Heartsome Translation Suite	Current build
255162675	Personal support from the developers.	DejaVu	7.0.0.284
255216913	Just coming up with text I'd had before but didn't remember.	DejaVu	Pro
255560075	The ability to search the entire TM for a term.	Wordfast	5.0zz2
256044514	Translation leveraging into updated projects.	Omega-T	1.6.0
256157535	concordance	Trados	6.5
256755918	That it usually works as intended. Btw, the previous question demanded an answer for the value for money question; this is not applicable because the tool is not sold.	LocStudio	7.3
257050929	MS Word integrated	Trados	
257089776	Single interface for all file types.	DejaVu	X
257223143	terminology adding/replacing in target text on the fly	MetaTexis	latest
257253646	Dictionaries that are made as you translate.	STAR Transit	XV
257524983	Helps to keep the same format as the original	Trados	7
257653907	concordance	SDL TRADOS 2006	
257682375	XML processing features	SDL TRADOS 2006	
258613923	Fast and easy-to-use	Omega-T	1.6
258638435	fuzzy matches/ align	Trados	6.5 lsp
258645771	CONCORDANCE	Trados	TRADOS 6.5 LSP

258672148	Term recognition	Trados	6.5
259028081	easy to use	Wordfisher	3.35X
259207737	The fact that it is a standalone program.	SDL TRADOS 2006	
259223735	Concordance	Trados	6.5
259232289	Concordance, Translate to fuzzy	Trados	6.5
259235006	saves time if there is a high rate of repetition	Trados	
259235580	Phrases or whole sentences that I have translated previously pop up automatically	DejaVu	DVX
259235864	Ease of entering terms into a glossary (highlight in both languages, press F11)	DejaVu	X
259237360	Excellent lay-out/arrangement throughout.	SDL TRADOS 2006	2006
259242168	The time it gives me to do something more interesting with my life!	Wordfast	
259246574	Concordance	Trados	7.1
259263059	The Word interface	Trados	7.1
259268748	The fact that it doesn't mess up Word formatting like Trados does.	Wordfast	
259273013	Price, very affordable and can do most things that Trados can do. Also compatibile with Trados.	Wordfast	
259293888	Concordance search	SDLX	2006
259311032	Good value for money	Wordfast	
259336064	Concordance - often saves chasing after terminology and often reminds me of meanings because it shows context.	Trados	6.5.5
259348129	The glossaries.	Wordfast	5.22
259356058	concordance	Trados	7
259358006	Ability to see work in Word environment.	Trados	7
259365587	User interface that melds with word processor.	SDL TRADOS 2006	7.5
259409985	Word interface	Trados	SDL TRADOS 2006
259417610	Layout of the translation page	DejaVu	7.0.284
259418681	I only use Star TM for Star UK jobs, so do not do any aligning. I would love to be able to use DV to align texts & build up a big memory.	STAR Transit	XVII
259453791	Concordance search	Trados	Freelance edition 6.5
259496917	Simple to use!	Wordfast	
259511792	'Concordance' function	Trados	5.5
259581468	consistency of terminology	Trados	

259586933	Automatic translation for 100% matches	Wordfast	9
259638382	Quite standard. It is easy to exchange TMs and terminology databases with colleagues.	SDL TRADOS 2006	7.5.0.756
259796742	repetition of same/similar phrases	Wordfast	
259938500	It's the only one that I have used and I am pretty happy with it - it just seems to work and doesn't do annoying things like crash.	Trados	6.5
259970317	The basic functions: presenting segments that I have already translated	Wordfast	5
259973706	It allows me to import all numbers, figures, email addresses, names etc. I make a lot of mistakes with these so my CAT tool helps reduce my error rate.	Wordfast	Sorry, I can't remember - it was the latest in November 2005
260327352	Being able to work in Word.	SDL TRADOS 2006	Latest
260349009	Cheap and easy to learn	Wordfast	
260351679	Bringing up my past translations	Trados	Freelance 7.5
260477159	concordance search	Trados	6.5
260657293	terminology search in documents	Trados	6.5
260752433	It's web-based.	My own	
260983339	AutoAssemble	DejaVu	7.0.284
261407039	Apply terminology	SDLX	2005
261410311	I love that it remembers what I've done in the past, and saves me having to hunt. The concordance feature is incredibly useful.	Wordfast	5.22
261415053	Heartsome runs natively on Linux :)	Heartsome Translation Suite	6.2
261416004	Analysis	SDL TRADOS 2006	7.1
261416820	Sharing memories, glossaries	DejaVu	
261422641	Simplicity, speed	SDL TRADOS 2006	FL 7.5.0.756
261425590	Easy to use after a short learning period	STAR Transit	XV
261430036	The concordance	SDLX	2005
261436472	Convenience	Trados	7.0 Freelance
261442479	The subsegment search and marking feature, especially for my marketing clients where the repetitions are mostly at the subsegment level.	MultiTrans	4
261460390	The improved way of using glossaries (with respect to the earlier version)	SDL TRADOS 2006	2006
261468901	Fuzzy search	SDL TRADOS 2006	7
261470673	Its easy to use	Wordfast	
261471057	Autosearch	DejaVu	3

261474351	The TagEditor! I like to see the 'bones' of a text.	Trados	5.5, but we'll upgrade to SDL Trados 2006 soon
261475871	easy to understand user interface (especially in Workbench), doesn't freak out new translators as they still can work in Word.	Trados	we have several versions in use, mostly 5, 6 and 7.1
261476192	Its improved structure and functionality over the earlier versions.	SDL TRADOS 2006	7.5
261476367	project filters	Trados	5
261479932	ease of use	SDLX	2005 SP1
261481620	Bitext generation	LogiTrans	Version 3
261481778	The full-text textbase that doesn't separate sentences from their context.	MultiTrans	4
261482128	Simple interface	DejaVu	dvx7
261482553	Concordance Search, TagEditor Preview, TM import/export	Trados	6.5
261483406	Easy to learn to use and easy to use	DejaVu	3
261483643	autopropagate	SDLX	2004
261485843	Definitely TermStar, the terminology management module (in which fields are customizable). I also really like that the user interface is totally customizable (in terms of colours, fonts, window size and location). The user can even decide which fields of	STAR Transit	Transit XV, service pack 18
261487955	s(t)olid reliability	SDL TRADOS 2006	7.5
261488512	The fact that I can use it in Word. The translation environment is familiar, so I know how to solve a problem if need arises.	SDL TRADOS 2006	SDL Trados 2006
261489348	concordance	Trados	7.1
261491464	concordance	Trados	6.5
261493140	Ability to process a wide range of file formats	SDLX	2005
261496465	The highlighted segments, they help me to concentrate better on the current sentence (and also the 100 matches of course :-)	Trados	7.0 (bought SDL Trados, but I'm still not using it)
261496575	Concordance search	Trados	6.5
261497816	Terminology management	Trados	6.5
261503273	Tag Editor Alignment	Trados	5.5
261510097	Word interface	Trados	5.5 freelance
261515232	More than one suggestion for fuzzy matches. Use of colours. Possibility to do real-time check against the glossary.	My company's internal one	
261518388	reversible bilingual glossaries	SDL TRADOS 2006	SDL TRADOS 2006
261524430	Context search	Wordfast	422 t22

261524456	Ability to work in familiar Word environment.	Trados	6.5
261531268	Interface with MS Word (so I can use Autotext and even macros, as well as spell check 'on the fly'.	Trados	6.5
261536811	FrameMaker Tagger	SDL TRADOS 2006	
261539866	Text analysis and interactive translation, plus terminology tool (Multiterm)	Trados	v. 7.1
261544299	Concordance search, Term recognition	Trados	7
261545063	Learnability and friendly interface.	Trados	6.5
261545944	Ease of TM management. Quality improvements in the latest version.	SDL TRADOS 2006	7.5
261546264	Concordance	Trados	6.5
261547815	Easy installation and updating, very simple way of creating a new project, easy to learn.	Wordfast	5.25
261549842	The context search	SDL TRADOS 2006	7.x
261550720	the ease with which terms can be added to the terminology module and the use of a pseudo-BTM (known as a VLTM) as secondary memory for both matches and context searches	Wordfast	5.25
261553881	All the filters to work with complex file formats	SDLX	2006
261558563	Its speed and ease of use	Wordfast	latest
261563470	Pretranslate	DejaVu	DVX 7.0.284
261566505	It is very easy to take over terminology data base entries using shortcuts Ctrl+1 ... Ctrl+9. (In Trados, the same procedure with Alt+Down is a real mess - slow and cumbersome).	SDLX	2004 Standard
261568131	The terminology database	DejaVu	DVX Pro
261571228	Working in Word	SDL TRADOS 2006	7.5
261573654	Segmentation	Wordfast	5.0h
261576023	Translator's Intuition can give us reusable text beyond the sentences. It does a great job with phrases and words.	Translator's Intuition	
261580371	Scan memory database	DejaVu	7.0.273
261585371	Having parallel texts.	SDL TRADOS 2006	2006
261589585	Full control over all stages of the translation	Trados	7
261596391	Ease of training	Fusion Translate CMT	
261620562	Content TM	SDL TRADOS 2006	2006
261647041	The capacity to work side-by-side (Source-Target) in the same window	DejaVu	DV X
261669116	Ease of use	Wordfast	
261670747	Keyboard shortcuts to enter glossary terms.	Wordfast	5.22b

261677308	concordance analyze link between TM tool and MultiTerm	Trados	6
261688936	Analyzing documents for leverage	SDL TRADOS 2006	2006
261702752	Analysis of files in order to find out the number of words/segments already available in the TM.	Trados	6.5
261747837	retreiving what it is stored in the memory	Trados	7
261808019	filter on selection - to display all segments that contain a highlighted term/expression - helps to make translation decisions	DejaVu	DV X build 284
261815883	Concordance	SDL TRADOS 2006	
261886885	Concordance	SDL TRADOS 2006	SDL Trados 7.5.0.756
261889230	Easy to use, extensibility	Wordfast	3.35
261898180	possibility to use iniform terminology	Trados	7.1
261903525	Concordance search	Trados	5.5
261906052	Being able to slot in the terminology from MultiTerm, and also to speed the process using the Word glossary entries, since the Word features are interoperable with Trados.	SDL TRADOS 2006	2006
261915360	check of repetitions	Trados	7
261915947	Workbench	SDL TRADOS 2006	latest
261923296	search feature for indiviudal terms within the sentences in the TM	Trados	6.5
261923462	Time savings	Omega-T	
261923716	For the certain projects, I can get very hight output per day/hour.	Trados	5.5
261949338	concordance	Trados	7.1
261952629	Creation of own file formats for import/possible modification of the available file format /customization	STAR Transit	XV
261992467	Time saved with big user manuals	Trados	6.5
262047792	Timesaving	Trados	7.1 Freelance
262226043	Concordance Direkt link to Multiterm Ease of use in Word	Trados	5.1
262271816	Analyse feature	Trados	6.5
262281437	The auto assemble tool that gives me a machine translation of the text. It is updated after each language pair. The next feature is the ease of adding terms and phrases to the databases... if it isn't really easy, it doesn't get done. This is crucial whe	DejaVu	the latest version DVX
262398984	Glossary.	Wordfast	5.22u1
262405130	The repetition rate in my line of work is close to O. I use LogiTerm to highlight similar terms/passages/sentences/blocks then quickly show me how I dealt with them last time or, occasionally, how someone else dealth with them. This is not, in fact, TM, b	LogiTerm	3.1

262549866	compatibility with many formats, convenient shortcut keys	Trados	Trados 6,5 Freelance
262641397	It works within Word and doesn't bother me with the interface that has to be studied.	Wordfast	
262657883	you can retrieve material from previous translations	Trados	6.5
262658106	Assemble	DejaVu	DVX
262670575	Interface (flexibility in segmentation, which can be changed at any point).	SDLX	Professional
262743131	Adding comments Selecting unfinished rows Propagating translations	DejaVu	7.0.384
262746672	Maintenance - > Filter can be used on both source and target languages, date, translator etc. And different critera can be combined.	Trados	6.5
262866939	Concordance search (retrieve old translations)	SDL TRADOS 2006	the newest, 2006
262888017	Able to compare documents	SDL TRADOS 2006	7.5
263115416	Clean up :-)	SDL TRADOS 2006	7.5.0.756
263115667	concordance	Trados	6.5
263242570	auto assemble function	DejaVu	DVX
263407201	It is translator oriented. This means it is translator-driven, relatively low-cost, constantly updated, and developed for translators instead of for corporate production. If I dream up a plausible feature that could improve my work and that of other tr	Wordfast	Always the latest. Wordfast is updated on a constant basis.
263652880	concordance	SDL TRADOS 2006	
263667620	efficient and intuitive user interface; ability to handle very large TM's	Memo Q	1
263704422	It's simple, easy to explain to others, probably the easiest tm tool for setting up a project, administration, easy to share tm, tb files... basically, the clearest gui so far	Memo Q	
263924516	texts are side by side, German on the left, English (Target) on the right, just as I would have them with hard copy	DejaVu	DV10
263924603	Easy to use, reliable and reasonable pricing, unlike Trados. I don't know why anyone would prefer Trados to WordFast, except that most big agencies require their translators to use it.	Wordfast	5
263951457	Ability to recall information	SDLX	
263957356	Segmentation	Wordfast	5.25f
263969931	Ease of use	SDLX	SDLX 2005
263971205	segmentation	Wordfast	5
263984183	Saves time	Trados	Trados 6.5 Professional
263989128	It works with Word	Trados	7.5
264023712	fuzzy terminology recognition	Wordfast	5.17z3
264193945	Automatic translation	SDL TRADOS 2006	7.5

264221395	The concordance tool.	Trados	5.5
264221975	Split screen with source language always visible. Color codes to indicate match value. Auto-propagate feature. Ability to 'lock' and filter material for faster translation.	SDLX	2005
264226968	export back into DTP software	SDLX	Elite
264229398	Wordfast is fully integrated with MSWord. It's also easy to integrate existing glossaries with it. Very simple to use. An excellent and economical TM tool.	Wordfast	5
264246517	I hope that one day, when I have more time, I will be able to utilise it better, having all those translations already set up with source and target sentences.	Wordfast	
264255774	To be able to apply the TM and have an approximate word count before starting the translation, so that I can book my schedule according to the work I'll really have to do.	SDLX	2006 (Build 6010)
264262072	I like the small things ... :) I love things like 'copy source with regard to lower/upper case' ... Things like that.	Wordfast	Wordfast 5.21i
264262451	That it is totally integrated into MS-Word.	Wordfast	5.1
264262995	Value for money	Wordfast	5,35 (the latest)
264263049	The ability to access to (a) three glossaries (b)three translation memories (c) one blacklist (e) untranslated reference materials.	Wordfast	Always the latest available
264263846	learnability, usability	Wordfast	the latest
264270187	All previously translated phrases at my fingertips	Trados	6.5
264276544	The simplicity	Wordfast	5.1
264326231	usability, and the chance to use non-native language in further translations	Trados	5.5, 6.5
264334537	context search	Wordfast	5.16e
264338454	Concordance	SDL TRADOS 2006	Freelance 2006
264351102	Terminology extraction	Wordfast	5
264352822	Efficient matches	Wordfast	4
264355107	terminology management	across	3.1
264377728	Dispaly (soure text above target)	Wordfast	5.17y2
264388793	The fact that you can translate a ligne/sentense at a time without seing the rest of the text. This make it easier and clearer to translate.	Trados	7
264403569	Consistency	Trados	6
264409231	Assemble	DejaVu	X
264414057	The 'translate to fuzzy' feature.	Trados	5.5
264419503	Terminology consistency	Trados	7.0 Freelance

264424981	The fact that previous translations provide me with a realiable bilingual database.	DejaVu	3.0.38
264431497	Concordance Tag Verifier	Trados	7
264441074	Concordance	Trados	6.5
264450248	Concordance	Trados	
264457376	Ability to work in Excel.	Trados	2007
264468203	The ease of building a glossary and accessing it.	Wordfisher	
264665587	usability	Wordfast	5.15|1
264666269	It's quick and easy.	Wordfast	
264666641	It automatically maintains a database of terminology	Wordfast	5.20n
264667148	Clarity of layout	SDLX	SDL Edit 2006
264668880	Hard to say, I like the whole user interface, its looks and the way it shows you the matches found in the memory.	Trados	7
264671251	exact or fuzzy matches	Trados	7.0 Freelance
264684870	All in all, I feel it helps me work better and faster. Cannot name any single feature, but I think I just could not work without Trados (or perhaps another TM tool).	Trados	7.1
264686496	It is quite easy to use.	Trados	6.5 Freelance
264699236	where to start? - ability to edit multiple files at once, ease of transferring terms from/to Termstar glossary, ease of management of reference material; range of features; speed.	STAR Transit	XV
264762421	Alignment	Trados	6.5
264774763	Concordance and win align	Trados	
264821160	Legacy text align.	SDL TRADOS 2006	
264853716	The concordance search since it allows me to find any thing I have translated before without having to read pages and pages of documents!	Trados	6.5
264866255	Interface	Trados	7
264985697	context lookup / term lookup	SDL TRADOS 2006	7.5 / 2006
265005380	Match recall	Wordfast	Beta version 07
265027173	Hard to say. Glossary	Wordfast	5
265073618	fuzzy matches, context search	Wordfast	5.24a
265076018	Replacing 100% or Fuzzy matches.	SDL TRADOS 2006	SDL Trados 2006
265123046	The 'translate to fuzzy' function and concordance search	Trados	Freelance 6.5
265128053	Concordance search	SDL TRADOS 2006	6.5

265144163	homogeneity	Memo Q	1.0.20
265234459	Auto-Assemble	DejaVu	
265238948	Translator's Workbench and Multiterm	Trados	7 - Freelace
265256147	The upgrades are free for 3 years after purchase.	Wordfast	5.22u1
265268212	concordance function	Trados	SDL X 7.x
265570349	Value for money	Wordfast	
265576581	Concordance search, terminology search	Trados	Trados Freelance 7.0
265873462	fuzzy matches	Wordfast	
265930314	Reliability, search functions	Clien'ts own tool	
265994777	Concordance Search	Trados	7
266214586	Aligning	Memo Q	1
266507194	Translate-to-fuzzy feature.	Trados	6.5
266761673	That I save time.	Trados	6.5
266828956	easy to install, easy to remove, easy to update, easy to maintain, easy to manage when working with several computers	Wordfast	4,1
267125548	The DejaVu 'assemble' feature	DejaVu	DV3
267265519	export,import of lexion, access to DB	DejaVu	3.07
267610732	Concordance	Trados	6.5.2 or 7
267697840	I am not fond of Transit at all. There is nothing I love there. Perhaps the only good thing is tahat it just skips all the repeated segments and takes you to the next one to translate. About Trados I like when it copies previuosly translated text and I f	STAR Transit	SP14
267721832	Advanced features, file format support, most common worldwide.	SDL TRADOS 2006	7.5
268236633	That it is a very goood way of not missig a sentence.	Trados	7
268336740	Efficiency in match: many repetitions in legal texts.	SDL TRADOS 2006	7.1.0.719
268828512	Lots of windows, but only one program (parts interact with each other seamlessly).	DejaVu	X
268883856	Segmentation	SDL TRADOS 2006	7.5
269258797	easy	MetaTexis	
269561977	Probably Concordance is the most useful when it works properly	Trados	6.5
271356559	The quality control tools	STAR Transit	Transit XV SP18
271551415	Reference search	Wordfast	5
273088959	Lexicon builder on the fly	DejaVu	DVX

273521449	easy to learn, easy to use, good support	Wordfast	5.22
273554600	saves a lot of time	Trados	7
273575995	Terminology Concordance	STAR Transit	XV
273584391	export	DejaVu	
273591316	CREATING MEMORY	SDL TRADOS 2006	
273618024	Corcondance feature Retrieving available matches	Trados	Trados 6.5 Freelance
273642191	WinAlign	Trados	6.5
273696683	Editing target and source text	Trados	6.5
273952390	concordance	SDL TRADOS 2006	7
276518725	TM is seperate from any particular project (i.e. you don't have to create 'project kits')	Trados	Trados 6.5 LSP / Freelance
280782135	Concordance	Trados	5.5
280789831	Reliability	Trados	7
280998119	Concordance	Trados	5.5
281063862	Its accuracy.	Trados	6.5
281076091	The 'Concordance' application.	Trados	6.5 Freelancer
281149213	concordance search	Trados	7
281163101	Usability	Trados	6.5
281410975	Functionality	Wordfast	
281468089	The user interface	SDL TRADOS 2006	7.5
281489191	Reliability, functionality	SDL TRADOS 2006	7
282117560	Concordance search	Trados	6.5
282279278	Concordance.	Trados	6.5 Freelance
284524907	Easiness of use.	Wordfast	5
284713506	concordance	Trados	5.5
284747232	Alignment, Analysis.	Trados	6.5 FL
284783367	analysis tool	Trados	7
285150990	The speed of it.	Trados	Trados 97
286938569	Concordance tool	SDL TRADOS 2006	6.5

Appendix E: Excerpts from TM Survey 2006 responses on desired features

RespondentID	Any comments for the above possible features? Is there any other particular task that you would like to see a Translation Memory do in the future?
253382438	No 2 sounds rather big brotherish to me. Am also worried about the fact that it would be difficult to validate the translation and/or source, apart from the fact that a lot of translations on the web are pretty awful.
253603039	And how would the copyright owner like it if I was to 'import dictionaries/ glossaries from a CD-ROM format'?
253646189	Well you should have asked about a certain data format to import dictionaries, but not CD-ROM format - a CD can have data in multiple formats
256044514	AFAIK, there's no such 'CD-ROM format' for glossaries. There're CSV and TBX, what do you mean?
253607643	Editing, and applying changes in TM, the translated text when reviewing it.
253622684	open api and python scripting possibility, open source the CAT tool, it should run on Linux
253628781	Semantic Sort out software
253638528	I don't really need the TM to automate terminology research for me: what I would value most would be the ability to import and export terms in multiple formats - e.g., XML, plain text, HTML for my own use but also whatever TM or database format the clients out there might supply or demand.
253646106	features are not that important. what matters is the workflow. right now all the workflows that are proposed are way too technical for the translator to fully appreciate the features.
253699691	Support for the Linux platform would be nice. Most TM suites force one to use Windows and Word for Windows. Many have suffered the consequences. The RTF file format in Word for instance keeps getting changed by MS, and companies like Trados (read SDL) can't keep up with these changes. Good luck with your survey, Peter
253700459	At the moment, the version of trados that I se can only find individual terms (termbase) or whole sentences (memory), it would be good if the memmory could store parts of senetences, phrases, as quite often part of a sentence will be a reprtition but this is not found by the memory. (This is where concordance is handy, as you can copy and paste the relevant bit into concordance to see how you transalted it before)
253700991	I don't think that a TM tool is supposed 'to bake a cake' while I am translating. There is a reason why it is called a translation memory tool and not machine translation. I believe that most of the things mentioned above are supposed to be handle by human translators. How on earth should a piece of software know whether or not a glossary located on the Web is accurate and can be trusted? Sometimes less can be more.
253703000	Just to work properly when connected in a network.
253704992	I interpreted 'very important' to mean 'oh, that would be cool' or 'DVX already does this and I find it very useful'. Stuff like PDF processing capability and an integrated OCR engine would require flawless text and format recognition, but I don't think that's feasible at the moment - I would never import a freshly OCR'ed document without reformatting and proofreading it first. Same goes for locating language resources on the net automatically - I trust myself more than a software in that respect, too.
253717435	More sophisticated example-based machine translation techniques, based on user-defined attributes on certain terms/phrases (part of speech, gender, etc).
253717539	The above 3 suggestions are very nice, but possibly a bit fancy. Can a TM tool do these any more easily than I can, and more importantly, should it? I'd prefer TM tools to concentrate on getting the basics right (why *can't* most others, apart from DejaVu, Assemble, and what good is a tool that doesn't, for example?) before they start adding loads of potentially useful but inessential additional features (more development time, more cost to the user, more things to update).
253724980	External spellchecking. No more insupportable limitation for number of used resources (TM, termbase, etc.) on the project. Support for open standarts like OpenDocument. New functions for translators, but not for programmers. Since 10 yers there are no new goog functions for translators (handling TM, termbase, concordances, term extraction, etc.).
253725113	These features may be useful - it's difficult to say without having tried them - but I'm getting on fine with one of the simplest tools on the market, which doesn't have any of them.
253734590	Combine machine translation features with translation memory features to make optimal use of both approaches.

253742851	I'd rather have a TM tool that is better and faster at TRANSLATING and doesn't do all these other things than one that does everything but slowly and poorly. I can google for bilingual texts, glossaries, etc.
253757061	improved formatting options as formatting can be tedious but is very much appreciated by clients
253757259	This is way above anything I ever dared imagine ... I'm not good enough to suggest even more useful improvements.
253765481	Our needs are quite simple, so it would be great if there was a 'pared-down' version of a TM tool that we could use.
253768493	The agencies need to be educated in the correct use of TM's, not apply it haphazardly.
253782436	I don't think the first feature is feasible due to copyright issues, but it sure would be nice.
253797296	have support for inflected languages, now it's a disaster to translate from Polish to English using any TM tool (stemming or lemmatising is required which is trivial to implement using dictionaries)
253816148	Some of these I would have to try out - completing the typing for example can be very irritating, like when Word suggests a spelling on the basis of a couple of letters and it isn't what you need. I do think the further development of online translation memory will be important, as two of my staff work from home.
253819300	I would love to see a TM tool have editing capabilities like EMACS.
253939769	Terminology propagation for new and updated terms Customized and sophisticated terminology handling features (current systems are too primitive) Better TM editing features and ease in their use by way of a more user-friendly editing interface (Trados is hopelessly complicated though quite powerful) Features to prevent corruption of the file formatting, and tools to identify/locate/fix any errors introduced by the tool Regular expression TM Search (yes, translators will be able to cope with regular expressions :-)) Interactive batch use of the TM (Trados has non-interactive batch use) directly from the word processor environment (more like programming environments in software development, such as Emacs)
253954652	Apply grammar rules to target text when putting together text segments
253996058	User definabale functions are needed, as Translation needs AI.
254015008	I tend to use other tools for OCR of 'soft' copies (FineReader), extracting PDFs (SolidConverterPDF) or making alignments (logiterm from Terminotix), so they are not priorities in the TM. But if it all came together on a neat package, excellent. In my wildest dreams, Atril (DejaVu) marries Terminotix! And gets rid of MS document dependance!! And runs on Macs!!! And PDF never existed!!!!
254016645	once the translation work completed, to fill a bilingual glossary in base of the source and the translated text.
254057940	I think there needs to be much more training offered for CAT tool users. I think that in the future the CAT should be used together with OCR and voice recognition to free us from keyboards and from sitting behind a desk all day. It would be great if there were a CAT program compatible with a mobile OS like Symbian, Palm or WindowsMobile/CE
254058528	If the features listed above would be available in one software solution, I would be willing to pay a premium price!
254079770	Can't think of any more - you've got some great ones here.
254085752	Point 2: I don't like Word with Trados because it can become very slow, but I like to see the context of the text. Point 3: I don't have a preference, performance is more important (i.e. with Trados, I like TagEditor better than TWB + Word, because the latter is slower, and both provide context info.)
254099994	The above possible features (question no. 4) strike me as quite impossible in most situations. Even if part of it can be implemented, most of it is inevitably doomed to failure due to its moving target shooting nature. That kind of work still had better be left to human beings. And all the effort these features would require would certainly be spent better if spent on the real critical features of CAT software.
254104382	Locate bilingual parallel texts on my subject on the Web and fill my TM database is a dream come true. Translate while I'm asleep, so that when I wake up the job is done and I can have a free day.
254113314	TM systems need to communicate/interact with their users better. If something goes wrong or if something is outside the capability of the software, error messages should be clear, simple, comprehensible and explanatory of the situation, so that the user understands what's going on and what caused the problem. Quite a few times, users have said: it did this and did that with no apparent reason. That's how users get frustrated and give up on the software. Error and other messages are crucial for the good collaboration between the user and the software.

254163472	Better documentation, advice and support for handling problems and errors. Trados is a very complex product with a huge range of options, and it is often very hard to understand the benefits of different options. This is not helped by the Trados documentation, which I find often rather impenetrable. I speak as someone who has written a number of technical user guides in the past, and I would do it quite differently from the way Trados do it -- for a start I would include far more example TMs to illustrate the various points. I would also include far more advice about how to recover from problems.
254163946	Integrated morphology ('low-level TM') could have been implemented years ago, but the developers don't seem to be interested
254175774	To be able to prioritize the Concordance matches by Client, Subject etc.
254180507	I would like to use the keyboard with easier key commands.
254187994	I'm more interested in getting free of Windows and Word than in having some kind of integrated front-end wrapped around a bunch of things that are by nature separate apps or distributed resources. A translator needs to mistrust any terminology from a remote or unknown source, so the source needs to be kept apparent when you start integrating remote terminology resources! By the way, why do you write 'hardcopy' when you mean 'print' or 'paper'?
254202383	all three make up a great wish list but if I can do it why couldn't the customer.
254283276	I'd like to be able to specify the order in which the programme lists suggested terms (i.e. assign an ordinal rather than only having the choice of chronological order or reverse chronological order).
254284689	wysiwyg match as you type
254591079	Those features sound interesting.
254615151	Smart autoassemble (using user-defined simple logic rules when retrieving terms from the terminology DB)
254668336	One full-fledged app please, but make it modular, so you can keep tasks apart when necessary, and with a *well documented and powerful* API, so that addons like e. g. language-specific assembly algorithms can be made and plugged in. I don't mind paying, but I do mind closed source lock-in effects and strongly resent platform lock-in.
254682410	Nothing more than the above
254682893	You say 'currently being developed' about 'construct and suggest fuzzy matches by combining different words/phrases (existing in your translation memory or in your terminology bank/lexicon), if an exact or fuzzy match cannot be found for a given source segment' -> that's my favorite feature in DejaVu. Even if it doesn't work 100% of the time it has already given me very satisfying results.
254714250	Streamline user interface (I am dreaming!).
254891740	Better 'terminology extraction', ie. extraction not based only on number of occurrences of a term in a text, but 'technicity' of the term (which may appear only once or twice but will require a lot of research to find the appropriate equivalent in the target language). The existing tools right now only make up lists of 'frequently found' terms, which 99% of the time are not technical terms related to a specific technical topic.
254954734	Integrated OCR and graphics would be wonderful. I don't know that I would want to *import* a CD-ROM dictionary, but being able to *access* one would be a huge plus.
255065232	Would like to see a text-only (Unicode) TM tool to work with OOo, say, or Enlaso.
256156997	Less expensive Term Extraction software
256309458	Allow not to use the tool for specific part of the text where it's not required or useful.
256755918	Is support for PDF really realistic? PDF was not designed to be translatable (or editable at all, apart from minor touchup) in the first place, and customers who send PDF for translation and don't want to pay for redoing the layout in a layout program are in need of training. Aligning previous translations, especially ones found on the web, is also not something I would do or recommend, apart from very rare cases like when a previous translated version of the same document is accessible.
257035458	Yes. When analyzing untranslated files, I would like to see an estimate of the time that will be required (based on an average value that I will enter) to translate them. What I mean is that 1000 words of a help file (eg. 100 strings of 10 words each) take less time than 1000 words of SW file (e.g. 500 strings of 2 words each).
257050929	more user-friendly interface and help
257089776	Integrated MT pre-translation for No matches.
257120158	The less clutter, the better. If a program is going to be looking for something on the Web, I want to be very sure that it is not going to tie me up with my ongoing translation.
259235006	The problem with TMs: agencies/clients will demand the use ot TM tools even when they are not suitable for the specific text; negative impact on payment for translators - translator has to invest in TM, learn how to use it and will get paid less oberall by the

	agency and get less work, too
259235347	these all sound very exciting and useful. Handling larger files without crashing would be good. Making coffee would be fab
259235864	Import scanned documents.
259252573	I don't understand the software solution question. Is it just referring to translation, or to translation business? The OCR capabilities and PDF capabilities are irrelevant if you have a good OCR program - don't know whether you ask about that later.
259253459	Not sure, haven't had the nerve to restart learning the TM I bought a few years ago. Have since switched to hard-copy based translation.
259263059	Be able to incorporate terms from a terminology recognition database more efficiently, like IBM's translation manager used to do: with a letter assigned to each term in the database which can be entered with 1 keystroke as you type.
259268748	All the features sound like good ideas in principle as long as they don't slow the whole process down and/or make it more prone to crashing or being irrational (e.g. replacing bullet points with pictures of post boxes - happened to me once when TagEditor got confused). Also, a cheap basic version without flashy features would always be welcome.
259358006	I am still using an old version of Multiterm because I have never been able to work out how to use Multiterm iX. Since the old version works fine, I haven't spent a lot of time on the subject.
259365587	A TM software program that would work just as easily in Linux and output MS Word compatible documents with the use of intermediate software. A programme that would deal with PDF documents directly without having to use OCR. A program that copes with Hebrew without changing character fonts and formatting.
259402759	All these features are unimportant. I don't use a TM, don't wish to use one...so why do you assume I use one above?
259409985	Improve reliabilty of WYSIWYG editing. Increase number of file formats that can be handled by WYSIWYG interface. 'Format painting' as used by SDLX is awful. Handle embedded graphics better (and more formats) would be a godsend.
259418681	I'm sure there is but can't think at the moment Have a spellchecker that is as good as Word
259581468	would be concrned about web searching being carried out by TM software, I think someone has a long way to go in that respect a) because of security issues and b) reliablity of texts/gloasaries on the Internet - forums win hands down in that respect
259638382	Yep. Some groupware features to enable me to work with colleagues on a current project would help create the ideal CAT tool.
259796742	recognise the differeence between 'dictionary' words and management rhubarb/invented words - very predominant in German. If I could ask my wife for her interpretation before starting a text then I could sail through the rest. (Perhaps the feature that the programme should have is a foreign husband or wife for all translators (preferably one who stays at home!)
259938500	I'm sure the link between the Internet and the CAT tool can be improved. It would also be great if Trados released a version for Windows Mobile 5 so I can take texts with me and edit them anywhere-but I think I'd be in a minority on that one. I'd like to look at Trados 7 as I think that has a lot of improvements in it. Otherwise, the software does all that I want - apart from the different word count structures that different companies use - but there are separate packages out there that allow you to do that. Good luck with the research!
260477159	improved fuzzy matches
261142915	I would expect a corpus of parallel texts would be useful if I was working in a particular field for any length of time but usually I am not.
261410311	handle edited texts after cleaning...
261422641	Do not disable any functions of the software for the source file (e.g. when translating .doc files, all MS Word features should be available) Having linguistic capabilites (e.g. recognizing singular-plural forms, inflextions, etc.) When, say, finding fuzzy matches from the TM, the ability to analyse and do some operations on the target segment as well.
261480621	HOW YOU ARE SURE THAT LANGUAGE PAIRS ON THE WEB ARE CORRECT? THEY ARE TRANSLATIONS...
261482128	Many web glossaries are suspect, duplicated or company-specific.
261487955	Improve sub-segment processing (as suggested above) - Incorporate MT at a user-defined level for sub-segment/new segment processing (e.g. search for terms in standard dictionaries) - Use multiple TMs/glossaries for secondary reference

261496465	check consistency of translation automatically
261496575	Implement client feedback automatically ;-)
261510980	What would be most helpful for me would be a search and replace plugin for Word with expanded functionality and integrated translation helps. I make much use of the search and replace facility in Word during translations.
261512936	Close the concordance box automatically after use.
261515232	Correct terminology extraction
261518388	Cross-referencing of glossary terms from Standards.
261531268	Shorter and less complex tags. More userfriendly and better explained setting options.
261539866	I would very much like to be able to centralize the content of all the dictionaries I have purchased and of glossaries available free on Internet, in order to query against them as I translate et have the possibility to insert translation of terms in my TM environment.
261545944	More format support.
261550720	I don't see why one would want to use a TM tool to search the web for appropriate texts, I'm sure the translator would be better equipped to do that.
261553881	More fuzzy matching capabilities at the term level
261563470	Make it easier to work with external translators, both those who use TM software and those who don't.
261566505	I wish the big players would include support for conjugated and declinated words in their terminology data base. In English a 'house' is always a 'house', so I can use a terminology base for that. In russian, it is 'dom', 'doma', 'domu', 'dome' and so on, depending on which case the word is in. The SDLX terminology base then just does not recognise the word. Trados Multiterm does some sort of fuzzy matching, but that one is unreliable. It might or might not find the word. The only tool that more or less handles this is Wordfast. There, you can enter 'dom*' and it will find all the variations. However, this could be improved on, e.g. 'dom*' also matches 'domashni', which has not got anything to do with 'dom'. So maybe a wildcard for just two or three letters would be a better choice, or even hand-crafted support for the case endings of certain major languages.
261585464	improved TM maintenance features, improved standards of interchangeability between TM tools, say between Trados and SDLX, more seamless interaction between Trados and SDLX. Improvement of TagEditor removing a variety of bugs etc.
261596391	Wow! How fantastic some of those features would be
261688936	In the feature list above, the one I consider THE feature of the future is the last one listed. TM tools need to go below sentence level and 'build' tentative sentences from bits and pieces found in TMs. This will boost productivity tremendously.
261815883	Automatic terminology building
261886885	Before adding new functions, Trados should take more care for a perfect functionality of their current products and for better more customer orientated support.
261889230	Supported should be: - Web-based time-tracking system accessible online to all localization teams on the project - separate software for manager (depending on needs) and freelancer (freeware) - localization testing tools, for example, tool to provide all possible screenshots for a localized applicalion (in pairs: before and after the localizatuion) to make it possible to perform visual testing of localization results on any screen
261906052	Terminology-management tasks performed conversationally within the translation environment (being able to edit MultiTerm from the Workbench environment so as to handle multiple equivalents for a source term, e.g., instead of editing the entries ex post facto).
261949338	would like availability of more than 1 reference memory for concordance searches, that can be sorted in order of priority
262047792	I've just finished working on a geological report (nearly 20,000 words rendered from Brazilian Portuguese into English) for which I had to use a huge amount of my time on the Internet (Google) to make sure the appropriate wording would be used (I am an aeronautical engineer, not a geologist!). If only I could have a user-friendly tool to help me in such situations...
262281437	Could I respond to this at a later date? I have an appointment now...
262398984	I am reasonably satisfied with what I have. I would like to see wordfast make the necessary changes so that I can use two different legacy fonts in the glossary. But I don't know if it is at all possible. I am not in favour of too many features. Translating with a TM tool is easier on the eyes and less taxing for my brain. That is enough reward for me.
262549866	If it could also do all the translating for me for the money it costs!!! I'd also include a more easily affordable price!
262670575	All of them are excellent options. They would help greatly.

263407201	Feature-rich does not mean overbearing. I would prefer a feature-rich, single small-scale tool that is highly-adaptable to individual translators' needs and--since most core TM features are already highly-developed in most TM tools--that places preference on new features that enable a translator to get started quickly on any format (OCR, PDF handling), to access additional reference materials more directly (CD dictionaries, major online dictionaries, bilingual parallel texts), and to work from within any of the major word processing applications and OSs. That being said, I don't believe that any one tool provider can develop a single end-all be-all tool.
263704422	I don't think that the TM tools should do the browsing on the net for me. I think that would be higly irresponsible anyway... however, once i find paralel corpora, it is a must that i can align it, and use it with the tm tool. i am not sure what you mean by 'cd-rom format' - if you mean, say, tmx, i think it would be essential to accept that (if not anything else) as an industry standard format.i find it highly frustrating that i need to obtain various tm tools for various clients - the interoperability between tm tools should is a basic interest of the industry - for end users and for tm tool manufacturers likewise.
263922013	I want something that can handle virtually all of the above. It's something of a pipedream, however, as I have never encountered software that is not buggy to the point of absolute frustration -- let alone software that handles the advanced functions described in your survey.
263924516	wow - it all sounds amazing. First I feel I have to become a 'power user' of what is already there in D,jaVu as I'm not using everything it offers already.
263944779	analyze the original file and highlight and reference all sections that are the same or similar.
263957356	Respect of original document formatting. XML compatibility to work with alternatives to Word and/or future Word format.
263976370	Better Help, I have learned TM myself and it is not very easy!
263985529	Mostly I found the automatic translation text have very poor grammer. I wonder if the latest T M tool can automatically correct and show you what should be the best grammer your translated text should use?
264173758	You will be having a bit of a problem with Farsi text as the font (parsnegar) is not fully developed yet and the recognition software is still being developed. But it would be very very noice to have a calender converter integrated as well. i.e. your GUI should include a small window for calender, a small window for exact word translation and a work area for typibng or text recognistion usable in both languages. Mind you you need to make it compatible with Farsi English texts put together. meaning one or two English words in the middle of a Farsi text. Cheers for thinking about this!
264193945	Maintain original formatting of the document
264221975	Extract terms from previous translations.
264226968	seamlessly integrate with a project managament system
264255774	To handle animations and text in images.
264262995	Suggesting the next words of the sentence as I type, possibly a few alternatives (e.g. 4-5) in a menu at the bottom of the page as I type.
264334537	I think integrating OCR is important on the source text input side and that integration with speech recognition is important on the translation input side. That is, to be able to dictate into the CAT program.
264336676	You have listed everything!
264379592	I would like to see some software that we can use between translators easily. I hope Trados could get PDF files normaly, that would make easy on the texts.
264403569	Allowing definition of segmentation, i.e., each unit would not necessarily correspond to a linguistic sentence.
264409231	allow the use of wildcards to reduce the number of entries in TMs and Terminology Databases
264419503	Quality Assessment functions
264441074	A reliable wordcount tool, includoing headers, footers, text boxes, text in images, etc.
264457376	I believe one of the most important features for translators would be voice recognition interface.
264583092	A well designed translation program should not be based on the mere mechanical replacing of texts (like the worst program in the world, the Trados) but it must think, analyze and reconstruct the text in the destination language, using its database to enhance the translation. I am using one of such programs, akin to the BabelFish, and I am quite satisfied. It allows you to work independently and to translate directly in Word (and in Word 6 as well).
264666641	This is the first time I have come across a mention of the TM tool actually *locating* webpages w/ glossaries etc. I very much doubt it will work, but if it did, it wd. be a boon.

264667148	Independency from MS Word features (e.g. independent spellcheck) could be an improvement.
264686496	It would be useful that there was a terminology consistency check tool integrated in the basic software packages; the tools themselves should be localized in more languages!
264699236	QA features for common mistakes, such as multiple spaces, spaces before full-stops, etc. Pretranslation analysis of documents for problem areas, such as in-sentence line breaks (Star has published stand-alone programs with this function for Word and FrameMaker files).
264985697	Be combined EASILY with an Translation Machine (Systran-like: maybe it does but I can't make it work with Systran 5 and Trados 6/7). I mean a feature integrated in TagEditor for example that would make it possible to run Systran on selected sentences that show no match at all.
266067320	Screen split horizontally and NOT vertically. Selection tabs along the top and not the sides.
266761673	Handle text boxes without problems
267287431	If it can have a programme were by we can store new terminologies for future references
267610732	Yes. It is very important to have an environment for the reviewer, for example identify only the segments to be reviewed and automatically open only these segments (e.g. sometimes a reviewer doesn't have to review 100% matches0.
267697840	Maybe all the above might be useful but as I don't work like that now i wouldn't know. I would like the software to be userfriendly, so you don't have to attend seminars to use them. And all those features they usually have but that normal users hardly ever come across the need for could be left out. The only thing that would be useful would be to be able to open all sorts of files easily and intuitively without haveing to look for buttons to click to get the right settings and translate them and get them back looking as they did before but in another language. And there preview is important.
267721832	I would like to be able to use Copy and Paste in an MS-Word environment. It is currently impossible to do so when using a TM tool.
268236633	Be compatible with other versions
268828512	Show concordances in a format similar to the corpus tools Wordsmith or monoconc/paraconc.
269561977	Would like it to work together with Wordfinder dictionary product (much used in Sweden) Would like it to handle InCopy documents and Lotus Notes produced documents
270157708	In order to appropriately and accurately leverage legacy translations, a TM tool should provide me with detailed contextual information on the suggested target text: filename, keywords, dates, folder location, and others I can't think of right now.
271356559	Include spellchecking in the terminology tool.
281063862	I think that the most essential are mentioned.
281149213	an ongoing count how many words/lines there are to go in the file or project, complete with match analysis
281489191	Better combination of fuzzy matches
282279278	Support Greek grammar (eg. pressing a key to get alternate versions of the translated word in plural, in a different inflexion).
285150990	Independent of Word program, clear and more simple
286303889	It depends how much work one is having.For small jobs one can not get much help

LONDON UNIV