Machine-Translatability and Post-Editing Effort: An Empirical Study using Translog and Choice Network Analysis

Sharon O'Brien

Thesis submitted for the Degree of Doctor of Philosophy

School of Applied Language and Intercultural Studies
Dublin City University

Supervisor: Dr. Dorothy Kenny

May 2006

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others save and to the extent that such work had been cited and acknowledged within the text of my work.

Signed: S. O'Brien

ID No: 91700981

Date: 30th May, 2006

This dissertation is dedicated to the memory of

Tuan Ó hAilín

ACKNOWLEDGEMENTS

There are many, many people, without whom this dissertation would have been impossible First and foremost, I would like to thank my supervisor, Dr Dorothy Kenny, for her unwavering enthusiasm, her excellent academic leadership, her friendship and support in all matters

I could not have embarked on this study without the financial support from the Irish Research Council for the Humanities and Social Sciences (IRCHSS) and the School of Applied Language and Intercultural Studies (SALIS) at Dublin City University

Equally, the research could not have been carried out without the support from IBM in New York and Stuttgart Dr Arendse Bernth embodied this support in New York and was responsible for securing an agreement which allowed me to use IBM's CL checker, EEA, their MT system, Websphere, and the user documentation for on ArborText Epic, an excerpt of which was used as the source text in this study I would like to thank Dr Bernth for proof-reading and offering very helpful suggestions on a draft of this dissertation I would also like to thank her for her foresight and determination Mr Ulrich Heess, of IBM in Stuttgart, also provided invaluable support by agreeing to allow twelve of his translators participate in this study To him, and to the subjects who agreed to participate, I say a big "thank-you"

I gratefully acknowledge the financial support of the Deutsche Akademischer Austauschdienst (DAAD) who awarded me a research travel grant This enabled me to spend an invaluable three weeks at the Institut der Gesellschaft zur Forderung der Angewandten Informationsforschung (IAI) in Saarbrucken where cutting-edge research and development on controlled language and machine translation is carried out In particular, I would like to thank Prof Dr Jorg Schutz for his support and guidance I would also like to thank Sun Microsystems for their permission to use their CL checker, Sunproof, which was developed by the IAI

I would also like to acknowledge the support of Arnt Lykke Jakobsen, of the Copenhagen Business School, who had the foresight to develop Translog and who granted permission for its use in this study

To Jeff Allen, for his enormous generosity in the sharing of information on controlled language and post-editing, and for his untiring promotion of these topics, thank you and please keep it up!

To my colleagues in SALIS, thank you for nurturing my interest in translation and technology and for all your support and encouragement

To my extended family, thank you for your encouragement and, especially, for the child-minding services when they were most needed!

But most of all, to Ruairi O hAilin – this dissertation most certainly would not have been completed without his technical genius, patience, love and support I am forever indebted Mile buiochas

Table of Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

AECMA	European Association of Aerospace Industries
CASE	Case's Clear and Simple English
CFE	Caterpillar Fundamental English
CL	Controlled Language
CMS	Content Management System
CTE	Caterpillar Technical English
DOCL	Dual-Oriented Controlled Language
EEA	EasyEnglishAnalyzer
GIFAS	Groupement des Industries Français Aeronautiques et Spatiales
GM	General Motors
HELP	Hyster's Easy Language Program
HOCL	Human-Oriented Controlled Language
IE	Information Element
ILSAM	International Language for Service and Maintenance
LTM	Long-Term Memory
MOCL	Machine-Oriented Controlled Language
MT	Machine Translation
NLP	Natural Language Processing
NP	Noun Phrase
NTI	Negative Translatability Indicator
PACE	Perkins Approved Clear English
PEP	Plain English Program
POS	Part of Speech
PP	Prepositional Phrase
QI	Quality Index
RTF	Rich Text Format
SAE	Society of Automotive Engineers
SDD	Siemens Dokumentationsdeutsch
SE	Simplified English
SECC	Simplified English Grammar and Style Checker/Corrector
STE	Simplified Technical English
SGML	Standard Generalized Mark-up Language
STM	Short-Term Memory
TCI	Translation Confidence Index
TM	Translation Memory

ABSTRACT

Studies on Controlled Language (CL) suggest that by removing features that are known to be problematic for MT (termed here “negative translatability indicators”, or “NTIs”), the MT output can be improved It is assumed that an improvement in the output will result in lower post-editing effort This study tests that assumption by measuring the technical, temporal and cognitive post-editing effort (Krings 2001) for English sentences in a user manual that have been translated into German using an MT system and that have been subsequently post-edited by nine professional translators The post-editing effort for sentences containing known NTIs is compared with the post-editing effort for sentences where all known NTIs have been removed In addition, relative post-editing effort (Krings 2001) – a comparison of post-editing effort and translation effort - is also measured The methodologies employed include the keyboard monitoring tool, Translog, and Choice Network Analysis Results indicate that temporal and technical effort is greater for those sentences that contain known NTIs Other findings are that some NTIs result in higher post-editing effort than others and that post-editing effort can also be expended for sentences where NTIs have been removed

INTRODUCTION

Lommel and Ray (2006) report that respondents to the LISA 2006 survey on Global Business Practices confirm that there is a growing diversity in the use of translation tools This suggests that, after a decade of use of translation memory tools, the translation industry is still engaged in a quest for better translation tools in order to automate the translation process even further (Van der Meer 2003) A number of technologies have increased automation and efficiency as their objectives – controlled authoring tools, alignment tools, translation memory tools, terminology extraction and management tools, software localisation tools, machine translation and content management systems The focus of this thesis is on controlled authoring and machine translation (MT)

The original idea for this dissertation was spawned when the researcher worked in the localisation industry (1993-1999) and observed that interest in MT and, subsequently in Controlled Language (CL), was on the increase Prior to and during this period, it was clear that an assumption was being made that the implementation of CL rules would improve MT output However, this assumption was being made on the basis of isolated examples or anecdotal evidence from practitioners who were usually not allowed, or who, for competitive reasons, did not want to share concrete data Even if this assumption proved to be true, another major assumption was then being made, namely that post-editing effort would subsequently be reduced No studies had been published that supported this assertion The aim of this dissertation is to address that question and, in so doing, to make an original and useful contribution to the field

Our first objective is to investigate whether or not the application of CL rules to a source text actually reduces post-editing effort Our second objective is to investigate whether specific CL rules have a greater or lesser impact on post-editing effort The study analyses data produced by nine professional translators acting as post-editors For comparative purposes, three additional subjects were asked to translate, rather than post-edit, the same text The source language is English and the target language is German and the subject domain is Information Technology The text used can be described as a user guide The machine translation system used to generate the target text is IBM's MT system, Websphere, which is used by IBM internally and is also available commercially

In Chapter 1, we discuss the historical background to CL, as well as more recent research The justification for, and benefits and drawbacks of CL are described A comparative analysis of eight CL rule sets for English is conducted and suggests that the implementers of CL/MT solutions are not in agreement on which linguistic features need to be "controlled" via CL rules

Chapter 2 focuses on the topic of CL for MT The main justification for implementing CL for MT, i e ambiguity reduction, is explored via a comparison of two specific CLs, Kant Controlled English and IBM's EEA This leads to a discussion of "translatability measurement", where the notion of negative translatability indicators is introduced We discuss four scholarly contributions to the field of translatability measurement, outlining the similarities and differences between models One of these models (Bernth and Gdaniec 2001) is selected as the guideline for the current study and a justification is given for this decision

The measurement of post-editing effort is one of the primary topics in this dissertation Chapter 3 introduces the topic of post-editing The objective of this chapter is to gain an understanding of what post-editing is and of what research has been carried out on the topic to date We explain the differences between post-editing, revision and translation We also look at the growing demand for post-editing services, as evidenced in publications over the last number of years Attention is drawn to the fact that the number of formal studies on post-editing is small The topics covered in studies typically include types of post-editing, post-editing rules, error classification, computer-aided post-editing and alternatives to post-editing The largest study on post-editing to date (Krings 2001) is introduced here We draw particular attention to one of Krings's assertions regarding post-editing, i e that an analysis of post-editing effort should include measures of temporal, technical and cognitive effort Finally, based on the assumption that the demand for post-editing services is on the increase, we look at the training and education needs of post-editors

Given that there are few formal studies of post-editing, Chapter 4 turns its attention to the related domain of Translation Process Research (TPR) The objective in this chapter is to draw on the methodologies used and results obtained in TPR in order to ascertain what might be useful in the study of post-editing We first give an overview of the main topics addressed by translation process researchers and we then consider what the implications might be for post-editing research One of the main methodologies used in TPR is think-aloud protocols (TAPs) Consideration is given to the pros and cons of this methodology in general and then more particularly in relation to the current research objective We also investigate alternative methodologies Consideration is given to keyboard logging, using the tool known as Translog, and to Choice Network Analysis A triangulation approach which employs both of the latter methodologies is deemed to be a satisfactory solution for the current research objective

Having made the decision on which methodologies should be employed, the path is paved for a more detailed examination of methodological issues in Chapter 5 The methodological framework is examined firstly from the theoretical viewpoint and then from the practical viewpoint We give consideration to the topics of independent and dependent variables, operationalisation, units of analysis and internal and external validity We then discuss the practical research design, addressing issues such as text type, number of words, language pair, number and type of subjects, the post-editing/translation assignment and data capture The empirical study at the heart of this research is then described and this includes the subject of data capture, where we describe in detail how we measure cognitive, temporal and technical post-editing effort

Chapter 6 presents the results from the data analysis We first report results on temporal effort, drawing comparisons between post-editing and translation times ("relative post-editing effort") We then turn our attention to technical effort where we examine the median number of deletions, insertions, cuts and pastes involved in post-editing In addition, we comment on patterns of dictionary usage as a measure of technical effort Finally, we examine cognitive effort We present data on pause ratios and comment in general here on the use of pauses as indicators of cognitive effort Using the data from Choice Network Analysis, we identify linguistic features that have a high, moderate or low impact on post-editing effort and draw some final conclusions on the correlations between post-editing effort and CL

In the final chapter, Chapter 7, we revisit the objectives identified here and ask whether or not they have been met Our findings, both anticipated and unanticipated, are highlighted We identify questions that could lead to future research and examine the methodologies used, the results obtained and the lessons learned

Chapter 1

1. CONTROLLED LANGUAGE

This chapter introduces the concept of "Controlled Language" (CL) and details its historical development from the 1930s through to the present Documentary evidence of the benefits and drawbacks of Controlled Language is discussed and the Chapter concludes with a detailed analysis of a number of controlled language rule sets [1]

1.1 THE DEFINITION OF CONTROLLED LANGUAGE

Huijsen (1998 2) defines Controlled Language as

> *an explicitly defined restriction of a natural language that specifies constraints on lexicon, grammar, and style*

He further divides Controlled Languages into two main categories, i e Human-Oriented Controlled Languages (HOCLs) and Machine-Oriented Controlled Languages (MOCLs) The objective of a HOCL is to improve comprehension by humans and that of a MOCL is to improve "comprehension" by a computer The difference between the two is described by Huijsen (ibid) in the following way

> *A general difference is that writing rules for the machine-oriented controlled languages must be precise and computationally tractable - for example, 'Do not use sentences of more than 20 words' -, while writing rules for human-oriented Controlled Languages may require skills that are presently beyond computers, and they may be somewhat vague - for example, 'Make your instructions as specific as possible' and 'Present new and complex information slowly'*

Of course, it may be intended that a Controlled Language should be destined for *both* human- and machine-processing We could apply the term "Dual-Oriented Controlled Language" (DOCL) in this case

Huijsen (ibid) also differentiates between "loosely" Controlled Language and "strictly" Controlled Language The former term applies when a CL is not precisely defined An example of this type of CL is *Perkins Approved Clear English* (or PACE) In contrast, a CL is defined as "strict" when it has a "formally specified syntax" (ibid 10) Huijsen alludes to the work of Cap Gemini's Lingware Services as an example of a strictly Controlled Language [2]

[1] Throughout this Chapter attention is drawn to the fact that few empirical studies on Controlled Language have been published In some sections in particular section 1 3 on the justification for Controlled Language I have drawn on my own personal experiences of working for five years in the Software Localisation industry as a language technology consultant During this time I had the opportunity to consult with companies such as General Motors and Sun Microsystems, to mention but two, on their authoring and translation processes Where sources of information are not given it can be taken that the opinions expressed stem from this experience

[2] At the time of writing Cap Gemini had ceased developing CL technology and offering consultancy

1.1.1 Sublanguage and Controlled Language

The term "sublanguage" was first used in 1968 by Zellig Harris (Harris 1968), who viewed sublanguage as a subset of natural language. Hirschmann and Sager (1982: 27) defined the term "sublanguage" as:

> *The particular language used in a body of texts dealing with a circumscribed subject area (often reports or articles on a technical speciality or science subfield), in which the authors of the documents share a common vocabulary and common habits of word usage.*

If we consider whether the term "Controlled Language" conforms to the above definition, the answer is: yes. A CL is a subset of natural language and is used by authors who share a common vocabulary and usage. What, therefore, is the difference between a "Sublanguage" and a "Controlled Language"? The answer lies both in the term "controlled" and in the word "habits" in the definition above. A sublanguage is a *naturally occurring subsystem* which develops over time in a community where people share common interests and communication habits. A CL, on the other hand, is an artificially constructed and controlled subsystem. It is true to say that CLs normally occur in communities where members share common interests and communication habits too, but their shared language is designed and controlled rather than naturally occurring. CLs and sublanguages differ not only in how they *occur*, but also in how they are *used*. A sublanguage can be both spoken and written whereas a CL is usually designed either for the production of written language or for the production of spoken language. It is possible to have a CL within a sublanguage, as Wojcik and Hoard (1996: 1) confirm:

> *Whereas formal written English applies to society as a whole, CLs apply to the specialized* **sublanguages** *of particular domains.* (emphasis added)

1.2 THE HISTORICAL BACKGROUND

The earliest known documentation of a CL is that of Basic English, which was developed by Charles K. Ogden in the 1930s (Ogden 1930). Basic English contained 850 words and some rules of inflection and derivation. Its aim was to create a foundation for the learning of standard English.

A well-known, early implementation of CL is Caterpillar Fundamental English (CFE). CFE is regarded as the predecessor of CLs whose aim it is to improve readability or translatability. The Caterpillar Corporation, who produce heavy machinery for large-scale construction work, introduced CFE in the 1970s. Kamprath et al. (1998: 52) list an increase in the number of products and in the labour costs as some of the reasons for the introduction of

CFE CFE had just under 1,000 terms, many of which had a broad semantic scope Although benefits were realised from its use (for example, there was a reduction in the number of technical enquiries arising from comprehension problems in the manuals (ibid 126)), CFE was discontinued after approximately 10 years The reasons for this were both technical and cultural CFE could not adequately describe the increasingly sophisticated products marketed by Caterpillar, Caterpillar could not enforce the basic guidelines, service technicians were not learning CFE due to limited education, and many cultures wanted their documents in their own languages

Although CFE was discontinued, it was used as the basis for other CL developments, most notably E N White's ILSAM (International Language for Service and Maintenance - (White 1980)), Smart's PEP (Plain English Program), J I Case's Clear and Simple English (CASE) and Perkin's Approved Clear English (PACE) (Douglas and Hurst 1996, Newton 1992, Huijsen 1998) Huijsen also credits Smart's PEP with giving birth to other CLs in turn, i e those used by Rockwell International and Hyster (Hyster's Easy Language Program – HELP), while ILSAM is reportedly the source for developments such as AECMA Simplified English (SE) (AeroSpace and Defence Industries Association of Europe - ASD), and CLs developed by Rank Xerox and Ericsson Telecommunications

During the 1980s and 1990s, research and development in Controlled Language was carried out at Wolfson College in Cambridge This R&D effort differs from the others mentioned above because it focuses on Controlled *spoken* Language Johnson (1996) reports on efforts to design and introduce a CL for police communication across the English Channel Similar efforts were made to develop a CL used for communication at sea, *Seaspeak* (Glover et al 1984), and for communication in the air, *Airspeak* (Robertson and Johnson 1988)

As mentioned earlier, Caterpillar abandoned CFE in the early 1980s However, developments in writing and publication technology allowed them to re-introduce CL in the form of Caterpillar Technical English (CTE) in 1991 Development of CTE happened in parallel with development of a CL Checker and Caterpillar's Machine Translation (MT) system (Kamprath et al 1998) CTE is still in use today and is seen as one of the most successful implementations of CL

AECMA Simplified English is another successful implementation of Controlled Language The requirement for safety in the aviation industry provided the impetus for the development and implementation of a CL Aircraft have become more complex and the documentation required to explain functionality has grown Farrington (1996) illustrates the exponential growth in aviation documentation by highlighting the fact that in 1910 approximately 100 pages were required to explain the functionality of an aircraft, while in 1990 500,000 words were required In addition, the need for clarity and accuracy in the manuals grew as the number of non-native speaking service technicians rose AECMA SE was officially renamed to ASD Simplified Technical English ™ in January 2005 [3]

AECMA SE was used as a basis for the development of other CLs in the aviation industry, for example, Boeing Technical English (BTE) (Wojcik and Holmback 1996) Wojcik et al (1998) explain that SE was too restrictive for the technical writers at Boeing Also, SE was not well-suited to non-procedural sections of maintenance documentation Hence, Boeing set out to define their own CL, based on SE, and to write a CL checker to support the technical writers [4] That CL checker, called "BSEC", is now available on a commercial basis to parties interested in implementing their own CLs

Simplified English served as the basis for the development of another CL, i e Français Rationalise (FR) FR was developed by "GIFAS" (Groupement des Industries Françaises Aeronautiques et Spatiales) Barthe (1998) explains how GIFAS was set up in 1985 with the aim of facilitating the introduction of SE into the French aerospace industry This group worked backwards from SE into French, assessing which rules were applicable to the French language, which were not applicable, and what new rules would be required They concluded that it was possible to design one CL to match another but that a "perfect match" was impossible (Barthe 1998 100) They also concluded that the texts written in FR were easier to translate because they were less ambiguous than texts where FR had not been applied

The automotive sector is another area where CL has been applied BMW and Volvo have both been involved in a CL R&D project entitled "Multidoc" (Haller 2000) Between 1995 and 2000, General Motors also invested in a CL project called CASL (Controlled Automotive Service Language) (Godden 1998, 2000) Indications are, however, that this project is no longer active [5]

With the exception of Français Rationalise, all the CLs mentioned above are based on the English language This reflects the fact that much documentation written for international usage is written in English However, there are numerous CL efforts for languages other than English For example, the German company Siemens has documented its efforts to produce a controlled German called Siemens Dokumentationsdeutsch (SDD) (Schachtl 1996) Schmidt-Wigger (1998) reports on grammar and style checking for German and Janssen et al (1996) report on Simplified German Efforts by the Swedish truck company, Scania, to

[3] See http //www simplifiedenglish aecma org/ for further information
[4] A CL Checker is a software application that applies pre defined CL rules to a document and highlights those parts of the document that break the rules Some CL checkers also suggest corrections to the user
[5] Personal communication Kurt Godden November 12 2001, reconfirmed on April 28, 2006

design and implement a controlled Swedish, which is then translated into several languages, are also documented (Almqvist and Sågvall Hein 1996, Sågvall Hein 1997, Sågvall Hein et al 1997, Sagvall Hein and Almquivst 2000) Cascales Ruiz and Sutcliffe (2003) also report on efforts to produce CL rules for Spanish

While heavy industries such as the aeronautic and automotive sectors have traditionally been at the forefront of CL research, design and implementation, there is now an increasing interest in the advantages of CL in other sectors such as the Information Technology (IT) sector Although not yet well documented, there is evidence that companies such as Sun Microsystems, Nortel and SAP have all recently invested effort in CL R&D [6]

As a final note in this section on the brief history of Controlled Language, it is worth noting two CLs which are different from all other CLs The first CL is known as Attempto Controlled English (or ACE) Attempto is a CL used specifically to translate software specifications into discourse representation structures and then into a logic language (Prolog in this case) (Fuchs and Schwitter 1995, Schwitter and Fuchs 1996, Fuchs and Schwitter 1996, Fuchs et al 1999) The second CL is called PENG (Processible ENGlish) (Schwitter et al 2003) Similar to Attempto, PENG is a computer-processible controlled natural language specifically designed to facilitate the writing of precise specifications (ibid 142)

1.3 The Justification for Controlled Language

1.3.1 Introduction

Currently, Controlled Languages are in use in many different domains, e g the automotive, aeronautic, heavy machinery and IT sectors There is a common general aim across these diverse domains, i e to produce quality information which is easily understood by the target audience and easily translated Hidden underneath these two general aims is a matrix of more complex considerations which motivate the design and use of Controlled Language With regard to the target audience, for example, consideration must be given to whether or not the primary target audience, defined here as those who read the documentation in the source language, are native speakers of the source language The level of education of the primary target audience and the tasks they must accomplish with the help of the information must also be taken into account

The second general goal of Controlled Language, that the information be *easily translatable*, also prompts questions about underlying objectives does *easily translatable*

[6] Sun Microsystems are listed as corporate sponsors for the Controlled Language Applications Workshop (CLAW) 2000, Nortel s implementation of CL is reported in Atwater (1998), the author has been personally involved in co supervision of a research project on CL for Symantec in Ireland (2005) and has consulted with Microsoft (Ireland) on the same topic (also 2005)

mean by computer or by human translation methods or a combination of both? And how does one measure the translation effort? Is it in terms of time, quality of output, post-editing effort, costs, or a combination of all of these factors? These questions are addressed in more detail below

1.3.2 Reported Benefits of Controlled Language

The field of Controlled Language research and implementation is relatively young and this is reflected in the fact that very few empirical studies have been published The majority of studies in the field of CL have been published in the proceedings of the three CLAW (Controlled Language Application Workshops) conferences (Adriaens et al 1996, Mitamura et al 1998, Adriaens et al 2000) and the joint EAMT/CLAW conference (EAMT/CLAW 2003) Despite the scarcity of published empirical studies, it is generally accepted that the use of Controlled Language can produce beneficial results

> *Although much work remains to be done on the evaluation of controlled languages before hard claims can be made, the few studies that are available support the belief that the use of human-oriented controlled languages improves the readability and comprehensibility, especially for complex texts and for non-native speakers*
>
> *(Húijsen 1998 12)*

Documented benefits can be classified into two main categories the first category centres around the *source* text and the second around the *target* text The key words associated with the first category include *readability* and *comprehensibility* as used above by Huijsen (1998)

1.3.2.1 READABILITY

The term “readability” is defined by Means and Godden (1996 107) as “a factor of the complexity of sentence structures, the amount of ambiguity, and the use of standardized vocabulary ” If a document has a high readability factor, then it is likely to have sentences with simple linguistic structures which are open to only one interpretation The obvious benefit of a high readability factor is increased comprehension This is of the utmost importance when the document’s aim is to instruct the user on how to perform certain tasks on a piece of equipment, for example, an aeroplane, in order to ensure the proper functioning of that equipment Ease of comprehension is especially important when the user is a non-native speaker of the source language Ultimately, document readability is linked to the preservation of human life As reported in Atwater (1998), there are a number of readability indicators, for example the Flesch-Kincaid, Coleman-Liau and Bormuth grades These readability measures count the number of syllables, words or characters per sentence and the length of the sentence

1.3.2.2 TRANSLATABILITY

Means and Godden (1996 107) link translatability to the readability factor of a document by stating that "translatability is a factor of the readability level, and the number of words " This does not give us a detailed understanding of what exactly "translatability" means, but Underwood and Jongejan clarify the concept, from the point of view of machine translation at least

> *The notion of translatability is based on so-called "translatability indicators" where the occurrence of such an indicator in the text is considered to have a negative effect on the quality of machine translation The fewer translatability indicators, the better suited the text is to translation using MT*
>
> *(Underwood and Jongejan 2001 363)*

The degree of intersection between CL-readability rules and CL-translatability rules will be discussed in more detail in the section on Rule Analysis in this chapter The notion of translatability, and machine translatability in particular, is central to this research Therefore, the concept will be discussed at greater length in Chapter 2

1.3.2.3 PRODUCTIVITY

When a CL is introduced with the aim of increasing readability and translatability, it sometimes has more far-reaching benefits In particular, when the translation process is automated using machine translation tools, the use of CL can improve *productivity* as reported by Caterpillar

> *Our experience thus far has demonstrated that CTE can have a significant positive impact on both authoring quality and translations productivity* [sic]
>
> *(Kamprath et al 1998 60)*

In today's business climate, where translations are expected to be available at the same time as source language documentation, increased translation productivity is a valuable benefit Note, however, that increased productivity is not associated with the authoring process In fact, the introduction of CL into the authoring process is likely to have the opposite effect This is especially so when authors use CL for the first time since they have to unlearn their existing writing habits and learn new ways of writing Additionally, authors have to master CL checking technology Tyson (1985) reports that research at Caterpillar estimated that the authoring process could be divided into different stages research, structure and writing *Research* accounted for 60% of the total time required for the writing cycle *Structure* and *writing* accounted for 20% each With the introduction of CFE, the 20% of time required for writing increased but soon returned to normal when authors became used to CFE

The fact that significant time is added to the authoring process is frequently seen as a drawback of CL However, for those companies who have been successful with CL, the increase in quality and translation productivity seem to negate this drawback because you write once – notwithstanding source language updates -, but translate into many languages

1.3.2.4 QUALITY

Using Controlled Language does not automatically guarantee a high level of quality in documentation Indeed, the concept of "quality" is particularly difficult to quantify Different organisations have different criteria for measuring quality For example, in the aeronautic industry, documentation "quality" could be measured by the number of clarifications required by a service technician before a maintenance task can be correctly executed on a piece of equipment In the automotive sector, it might be measured according to the number and type of errors made according to the Society of Automotive Engineer's (SAE) translation quality standard – the J2450 (Woyde 2001) In comparison to this, the "quality" of a software manual could be measured according to the ease with which required information can be found and followed (Byrne 2004) Generally, Controlled Languages are tailored towards the specific goals of the user In the case of General Motors (GM), for example, the aims were increased quality of source documentation, leading to better service and, ultimately, lower costs, but there was the added benefit of complying with language laws in certain regions

> *In addition to cost savings, we have high confidence in predicting the following additional benefits for GM in the U S as well as worldwide - improved quality of source documents, - increased usage of service manuals by technicians, - improved ability to service vehicles correctly the first time, - lower warranty costs, - higher customer satisfaction, - increased sales Pertaining specifically to translation of service manuals, the following additional benefits are expected - reduced lead time in producing translations, - compliance with language laws in Quebec and other countries*
>
> *(Means and Godden 1996 109)*

1.3.2.5 RETRIEVABILITY, RECYCLING, REUSE

Whether an organisation is trading in sectors as diverse as the heavy machinery or IT sector, the challenges of commercial survival are similar meet your customers' demands for new and better functionality or they will purchase elsewhere This translates into a requirement for new or updated models or versions on a frequent basis When a new feature is introduced, it must be documented and the information must then be translated However, this information is not distributed in isolation but is added to existing information and entire manuals, be they hard-copy or electronic versions, are re-published Therefore, retrievability of previously published information becomes important

When an author is required to update a manual, there is a temptation to "improve" sections written previously Authors who work in mono-lingual (usually English-speaking) environments are frequently unaware of the cost of these improvements Every punctuation mark or word changed results in increased translation costs and, the more target languages the information is due to be published in, the higher the costs Therefore, if authors were trained to re-cycle previous information without making unnecessary changes, one could justifiably expect a reduction in the cost of translation

Recycling previous translations is also recommended In the translation industry, re-use has practically become the norm in some sectors as a result of the implementation of translation memory technology Re-use of source information, on the other hand, is only slowly becoming the practice Source information is best re-cycled using Content Management technologies Content Management Systems (CMS) store data in the form of "Information Elements" (IEs) An Information Element is a stand-alone chunk of text In traditional writing terminology, it could be a section, a paragraph, or even a chapter Most commercial Content Management Systems store IEs in SGML format Retrievability and maintainability of information is the main selling point of these systems

If an organisation has invested in Controlled Language, then the incentive to re-use, rather than re-create, information is even stronger Although *retrievability* is not a direct benefit of CL it can be included as one of the objectives in introducing CL

> *The objective of a CL is to improve the consistency, readability, translatability, and retrievability of information*
>
> *(Wojcik and Hoard 1996 1)*

In the high-volume translation business, retrievability and reusability of translations is a must Thus most high-volume translation companies have implemented Translation Memory systems (TMs) (O'Brien 1998) The combination of Content Management Systems with TM, CL and MT presents a powerful solution for the retrievability and recycling of information and is a solution that is being implemented by some companies (O'Brien 1999)

1.3.2.6 Cost Benefits

While readability, quality, and productivity are frequently listed as the benefits of CL, reduction in the cost of producing multi-lingual information is undoubtedly one of the most sought-after benefits As stated previously, few empirical studies on CL have been published This can be attributed to the fact that the implementation of CL is most often executed in a proprietary environment Thus, there is little published evidence that using a CL reduces information production costs

Nevertheless, we can deduce that some cost reduction is possible since CLs are often implemented *in conjunction with* machine translation and the two main objectives of MT are to reduce translation time and cost It is well known in the translation industry that the rate of pay for post-editing MT output is substantially lower than that of human translation

Douglas and Hurst (1996 94) confirm that cost reduction is possible when CL is twinned with MT technology

Perkins report a cut in translation costs of between 50% and 70% with the use of PACE

Despite this confirmation and the anecdotal evidence that costs can be reduced, to my knowledge, nobody has yet publicly documented how much it costs to introduce and maintain CL in an organisation and how these costs are offset against the benefits of CL

1.3.3 Reported Drawbacks of Controlled Language

It is inevitable that the introduction of a tightly controlled writing and translation environment will have its drawbacks as well as its benefits The main drawbacks are explored in this section Publications on the topic of CL tend to focus on the challenges and benefits of implementing CL Few, if any, organisations will publish information on their *failures* with CL

1.3.3.1 Costs

Undoubtedly, the costs of implementing a CL act as a disincentive to organisations and may also be the reason why some recent CL efforts have stopped or been put on hold While actual costs will vary depending on the economic environment and the effort expended, we can discuss potential costs by examining the tasks required in the setting up of a Controlled Language These stages can be listed as follows

Definition and Creation of a CL

Development or Customisation of the CL and associated technology

Training

Maintenance

In the first stage of definition and creation, a team of researchers is required to examine the suitability of existing CL definitions and to define the CL requirements of the organisation in question The second stage involves either development of a new CL or customisation of an existing CL This stage also involves the creation of a controlled terminology base and either the customisation of a commercial CL checking tool or the development of a proprietary tool Training of authors in the use of the CL rules and the CL checker is required in the third stage Maintenance is the last stage Maintenance is an ongoing effort whereby feedback on rules and terminology and checker efficiency is accepted and changes are implemented accordingly During this stage, evaluation of the efficiency of the process should also be carried out

The number of highly specialised human resources required for each of these stages is significant Understandably, any organisation considering the introduction of CL has to be convinced of the benefits before embarking on such a time- and resource-demanding project

1.3.3.2 User Resistance

The user group most likely to be affected by Controlled Language is authors While technical authors can be trained professionally, it is not unusual for a technical author to first be trained as a software or civil engineer, for example, and then as a technical writer This is significant for the introduction of Controlled Language because the user group might not have any formal linguistic training and awareness of translation requirements can be limited Introducing Controlled Language can be viewed as a threat to the very people who must embrace it Successful implementation of CL in such an environment could be an unobtainable goal

Authors view CL as a threat because they are being asked to unlearn their writing habits and to relearn new, restricted ways of writing Writing using a Controlled Language is particularly difficult since demands of conformity and consistency are put on the writer Also, new technology, in the form of a CL checker, must be learned and applied Nevertheless, user resistance can be overcome Reuther (1998 180), reporting on the degree of scepticism and the fear of restriction displayed by authors who were being asked to use new CL technology, explains how this fear was overcome

> *the technical authors became aware that the tools they are intended to use will not dictate to them in an inflexible and arbitrary way a certain style of writing, but that the tools support them by merely giving indications where possibly a correction or reformulation might contribute to more clarity and readability in their documents*

Clearly, flexibility in the technology contributes to reducing user resistance

Another way of reducing resistance is by involving the authors in the consulting project from the very beginning and accommodating their existing work practices where possible Interestingly, General Motors seemed to learn this over a period of time when introducing CASL into their organisation In 1996 their intention was that CASL

> *will impose new restrictions on the work of authors and translators, and will drive some changes in their work patterns*
>
> *(Means and Godden 1996 110)*

However, by 2000, their approach had changed significantly

> *We realized that it was essential for success to minimize the impact of CASL technology on the authoring community*
>
> *(Godden 2000 15)*

1.3.3.3 Complex Production Cycle

Introducing CL into the information production cycle means lengthening the cycle and making it more complex by adding the extra steps of CL checking and revision Checking can be carried out *after* an initial draft has been written or *during* the writing process The checking process can be carried out by the authors themselves or by a dedicated human CL editor The advantage of using a dedicated editor is that the impact on the author's writing processes is reduced Authors simply have to implement the CL editor's corrections The disadvantage of this approach is that the authors have less incentive to learn the CL rules and are more removed from the technology Also, having numerous authors and only one or two CL editors means that a human bottle-neck can exist

Godden (2000) reports on the approach GM took when introducing CL Their guiding principle was one of "minimal disruption" (Godden 2000 19) They tested three different models the author-centric model, where the author performed the CL checking and had to know the CL rules intimately, the editor-centric model, where the authors had no responsibility for CL and CL editors had to check documents and implement rules, and, finally, the hybrid model, where authors performed a once-off check and editors performed a more thorough check GM finally settled for the hybrid model The reasons for this decision are discussed in detail in Godden (2000 14-18)

Changes in existing processes, especially when this means additional steps, are not easily accepted in organisations Once again, the key seems to be to involve those concerned from an early stage so that they accept ownership of the new processes

1.3.3.4 Integration with Machine Translation

As mentioned previously, CL is often introduced at the same time as Machine Translation As so many examples have shown (Adriaens et al 1996, Mitamura et al 1998, Adriaens et al 2000, EAMT/CLAW 2003), the quality of MT output can be significantly improved when the input is controlled However, introducing Machine Translation technology also comes at a very high price Not only are the licences for MT expensive, but also time and resources are required at different stages of implementation, such as evaluation, development of rules and terminology, training and maintenance The link between Controlled Language and Machine Translation will be discussed in more detail in Chapter 2

1.4 CL Rule Analysis

Thus far we have seen that CLs have been implemented by a number of organisations, and we have discussed their purported benefits and drawbacks In this section we take a detailed look at individual CL rule sets in an effort to generalise about how CLs aim to improve translatability To this end, the author compared eight CL rule sets for English The results are reported in O'Brien (2003) and we will summarise the findings here

The eight Controlled English rule sets included in the analysis are AECMA STE, Attempto Controlled English, Alcatel's COGRAM, IBM's EEA, GM's CASL, Oce's Controlled English, Sun Microsystem's Controlled English and Avaya's Controlled English [7] Since CLs are often seen as proprietary, the decision to include specific CLs was primarily influenced by the willingness of developers to give permission for their CL to be included in the analysis Section **Error! Reference source not found** gives a brief description of each of these CLs

1.4.1 Description of the Controlled Languages included in the Analysis

1.4.1.1 AECMA Simplified Technical English

The AECMA SE Guide was first released in 1986 As mentioned previously, it has recently been renamed "STE" The release that was available to the author for this analysis was Issue 1, Revision 2, which dates from January 15th, 2001 Therefore, reference is made throughout to "AECMA SE" and not to STE

It could be argued that a comparison of Controlled Language rule sets should be confined either to Human-Oriented Controlled Languages or to Machine-Oriented Controlled Languages so as to compare like with like Nevertheless, AECMA SE, which is characterised as a HOCL, is included in this comparison, for a number of reasons

1 AECMA SE provides us with one of the few successful cases of a Controlled Language in use

2 The SE rules are constantly revised by a panel of experts and the manual is a controlled document which is available publicly

3 AECMA SE has been very influential in the Controlled Language domain and there are numerous articles on the subject (e g Kincaid 1997, Shubert et al 1995, Adriaens 1994, Hoard et al 1992) Some CLs are derived from AECMA SE (e g Boeing

[7] EEA stands for EasyEnglishAnalyzer" To avoid confusion with other Easy Englishes the distinction that used to exist between the CL EasyEnglish and the CL checker EasyEnglishAnalyzer has been dropped and both are now referred to by the acronym EEA (Personal Communication Dr Arendse Bernth January 2006)

Technical English) Also, one of the commercial CL checkers (MAXit) checks for adherence to SE rules

4 Acquiring CL rule sets is not an easy task and it was judged that it would be more fruitful to include all rule sets possible in the analysis rather than to leave one out because it was classified as a HOCL and not an MOCL like many of the others analysed here

In addition, although it is classified as a HOCL, there is reason to believe that some SE rules can be equally applied to an MOCL Reuther (2003) reports on an analysis where rules for improving readability in German (called "R-Rules") were compared with rules for improving translatability ("T-Rules") R-Rules equate roughly with rule sets for HOCLs, as described above, and T-Rules equate with rule sets pertaining to MOCLs Reuther concludes that "readability rules are a subset of translatability rules" (ibid 131), thereby suggesting that all rules governing readability also improve translatability Reuther, however, does not provide concrete definitions of readability and translatability

1.4.1.2 ATTEMPTO CONTROLLED ENGLISH

Attempto Controlled English (ACE) was developed at the Computer Science Institute ("Institut der Informatik") at the University of Zurich (Fuchs and Schwitter 1995, 1996, Schwitter and Fuchs 1996, Fuchs et al 1999, Schwertel 2000) As indicated in section 1 2, Attempto is a CL used specifically to translate software specifications into discourse representation structures and then into the logic-based programming language Prolog The objective, according to Schwitter and Fuchs (1996 3), is to "improve the quality of specifications without losing their readability, ( ) to restrict natural language to a controlled subset with a well-defined syntax and semantics that can serve as a suitable view of a logic language " Attempto, along with the later development, PENG, are the only CLs known to the author which "translate" a natural language CL into an artificial language This makes Attempto an interesting candidate for inclusion in the comparison of CL rule sets ACE has two different types of rules construction principles and interpretation principles Construction principles dictate what linguistic features can or cannot be used, e g *Do not omit the relative pronoun "who", "which" or "that"* Interpretation principles dictate how certain linguistic constructions are to be interpreted by the system that uses ACE as input, e g "*A personal pronoun always refers to the most recent accessible noun phrase that has the same number and gender*"

At first glance, classifying ACE as either a HOCL or an MOCL is problematic On the one hand, it is a HOCL because it provides rules to improve the quality and maintain the readability of specifications created by humans On the other hand, it can be classified as an

MOCL because its aim is to translate software specifications from natural language into an artificial language It is the author's opinion that ACE has more to do with machine processes than with human-oriented readability I have therefore chosen to classify ACE as an MOCL, rather than as a HOCL

1.4.1.3 Controlled Automotive Service Language (CASL)

CASL was developed between 1995 and 2000 by General Motors with the aim of reducing the cost and time involved in translation and to improve the general readability of their documentation (Means and Godden 1996, Godden 1998, Godden 2000) CASL can be classified as an MOCL since its primary objective is to reduce the cost and time of translation using a combination of authoring and machine-assisted translation technology

1.4.1.4 Cogram

Cogram, the CL designed and used at one time by Alcatel, is described by Schreurs and Adriaens (1992) and Adriaens (1994) Cogram was the underlying CL grammar used in the LRE2-funded SECC project (Adriaens 1994) [8] SECC stands for "Simplified English Grammar and Style Checker/Corrector" According to Adriaens (ibid), Cogram has a total of 150 grammar rules which are organised into the following four major categories textual control, syntactic control, lexical control and character and punctuation control In addition to the grammar rules, there is an English lexicon of 1,500 words ("Colex") and a restricted English lexicon containing words in the domain of telephony ("Cotech") Alcatel is no longer using Cogram [9] The SECC project treated CL checking as a specific MT problem an MT engine (METAL) was used to convert the source text into a SE compliant version Hence, Cogram is classified as an MOCL

1.4.1.5 EEA

A description of IBM's EEA can be found in numerous articles, e g Bernth (1997, 1998a, 1998b, 1999a, 1999b, 2000) The tool which checks for compliance with EEA is described by Bernth (1997) as a "grammar checker++" because it combines grammar checking with CL checking Approximately 40 checks can be carried out by the checking tool EEA is used to improve the translatability of documents by MT and can therefore be classified as an MOCL

[8] An EU-funded programme which financed projects in the Language Engineering domain

[9] Personal communication from Patrick Goyvaerts, independent consultant to, and former employee of Alcatel, on 14 March 2002

1.4.1.6 Océ Controlled English

Oce is a Dutch printer manufacturer Oce's implementation of Controlled English is mentioned in Cremers (2001) Through personal communication with those responsible for the CL effort within Oce, it was possible to establish that Oce uses Controlled Language to improve the translatability of documents by machine (specifically, the Logos MT system), and that the Smart MAXit checker is used to check compliance [10] The MAXit checker checks for compliance with AECMA Simplified English, but can be customised for other CLs

Oce customised only two rules The first customisation was an elimination of the rule that requires numbers under ten to be written in words [11] The second customisation was the addition of a rule called the "Dutch Glue" rule Native Dutch speakers have a tendency to write compounds in English as they are written in Dutch, i e as one lexical unit This rule checks for such occurrences (Lou Cremers, personal communication, March 2002)

Given that Oce currently translates documents using MT, it would seem logical to classify the Oce CL as an MOCL However, it should be noted that the majority of rules for this CL are rules derived from AECMA Simplified English, which is categorised as a HOCL

1.4.1.7 Sun Microsystem's Controlled English

Sun Microsystems is involved in experimental efforts with Controlled Language and Machine Translation Sun Microsystems is interested in CL because it can improve translatability, in particular, machine translatability, and can potentially reduce the time required for translation For this reason, Sun Microsystems' CL is classified as an MOCL

1.4.1.8 Avaya's Controlled English

There are no published articles to date describing Avaya's Controlled Language research and implementation Avaya developed its own CL and used this to improve source documentation, human and machine translation (personal communication, Jane Lynam, 2002) It is therefore classified as an MOCL

The rule sets for the following CLs were obtained on condition that confidentiality be maintained CASL, EEA, Sun Microsystems' CL, Avaya's CL For this reason it is not possible to reproduce these CL rules here Although this placed some restrictions on the analysis that was carried out, it was still possible to report on the phenomena the rules govern, the types of rules, and their frequency Where comments are made in the next

[10] When Océ changed from authoring in Word and FrameMaker to XML, they experienced problems with their CL implementation with the result that the number of people using CL in the organisation dropped (Personal communication Lou Cremers, February 2006)

[11] The reason for this is that Océ also employs translation memory technology If they had implemented this rule, they would have lost a number of exact matches in their translation memories and this in turn would have cost a lot of money Therefore, they decided to eliminate this rule

section on specific rules, the CL to which this rule belongs is only mentioned explicitly if that CL is not subject to a confidentiality agreement

1.4.2 Rule Classification

The difficulties associated with classifying CL rules are discussed in O'Brien (2003) For the analysis reported here, rules were classified according to the linguistic feature that the rule sought to control The three categories were

- Lexical, with a sub-category of lexical-semantic
- Syntactic
- Textual, with two sub-categories of Text Structure and Pragmatic

AECMA SE rule 3 1 is an example of a lexical rule Use only those forms of the verb that are listed in the Dictionary SE rule 1 3 is an example of a lexical-semantic rule Keep to the approved meaning of the word in the Dictionary Do not use the word with any other meaning SE rule 2 1 is an example of a syntactic rule Do not make noun clusters of more than three nouns Rule 4 3 is a text structure rule (Use a tabular layout (vertical layout) for complex texts), while rule 6 2 (Try to vary sentence lengths and constructions to keep the text interesting) is an example of a pragmatic rule [12]

1.4.3 Number of Rules

For the reasons outlined earlier pertaining to confidentiality, six of the rule sets are referred to as CL1, CL2 and so on up to CL6 Table 1 1 shows the total number of rules for each CL in the analysis

Controlled Language	*Number of Rules*
AECMA SE	60
ACE	36
CL 1	59
CL 2	46
CL 3	35
CL 4	31
CL 5	36
CL 6	38

Table 1 1 Number of Rules in Each CL

Table 1 2 shows the percentage and number (in brackets) of types of rules in each CL, i e Lexical/Lexical Semantic, Syntactic, and Text Structure/Pragmatic

[12] The classification of rules is not without problems This issue is discussed in more detail in O Brien (2003)

CL	Lexical / Lexical Semantic	Syntactic	Text Structure / Pragmatic
AECMA	(18) 30%	(12) 20%	(30) 50%
ACE	(11) 30%	(23) 65%	(2) 5%
CL 1	(26) 45%	(24) 40%	(9) 15%
CL 2	(9) 20%	(32) 70%	(5) 10%
CL 3	(15) 45%	(11) 34%	(7) 21%
CL 4	(7) 22%	(13) 42%	(11) 36%
CL 5	(9) 25%	(18) 50%	(9) 25%
CL 6	(17) 45%	(15) 40%	(6) 15%

Table 1 2 Number of Types of Rules in Each CL

The numbers in Table 1 2 are represented in visual format in Figure 1 1

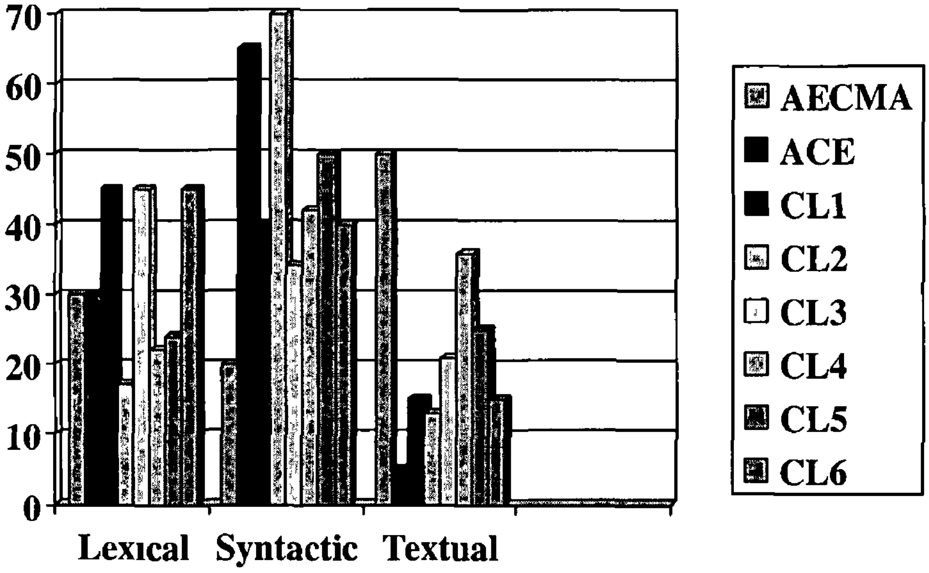


Figure 1 1 Number of Types of Rules in Each CL

Some general observations can be drawn from the table and chart above Syntactic and Lexical rules account for the largest proportion of rules overall in the group of CLs analysed With the exception of AECMA SE, text structure and pragmatic rules makes up the smallest category for each CL The number of text structure/pragmatic rules is low because pragmatic rules govern text function and CL checking technology is not capable of deciphering text function [13] It is interesting to note that AECMA SE, the one CL characterised as a HOCL in this analysis, has the highest percentage of textual rules

[13] The use of SGML tags to identify the function of a sub text is of course, possible and some efforts have been made to make use of SGML tag checking capabilities in CL checkers (e g Mitamura and Nyberg 1995)

AECMA SE and CL 4 have a higher percentage of text structure rules than any of the other CLs An analysis of the eleven text structure rules in CL 4 reveals that only three of these rules are shared with AECMA SE The remaining eight are unique to CL 4 and focus primarily on punctuation governing, for example, the use of exclamation marks, semi-colons, parentheses etc , whereas text structure rules in SE focus more on *information structure* and information load than on punctuation

The percentage of syntactic rules included in the AECMA SE rule set is considerably lower than in all other CLs (i e 20% versus 34%-70% for the other CLs) Finally, CL 2 has a noticeably lower percentage of lexical rules built into the rule set (i e 20%) in comparison with other CLs (the highest percentage of which is 45% for CL 1) It is worth pointing out that CL2 also has the highest proportion of syntactic rules, i e 70%

1.4.4 Shared Rules

It is remarkable to note that only one rule is common to all eight CLs under comparison SE rule 5 1 "*Keep procedural sentences as short as possible (20 words maximum)*" is echoed in different ways by all CLs where the maximum number of words allowed in a sentence varies from 20, for instructional sentences, to 25 for descriptive ones Other CLs simply urge the writer not to be too verbose

1.4.5 Common Rules

"Common Rules" are defined here as rules that are shared by at least four (i e 50%) of the CLs under analysis The following list details the rules shared by four or more CLs

- SE rule 1 1 "*Use approved words from the Dictionary etc*" is shared by three other CLs While a controlled lexicon is as important in a Controlled Language as the rules themselves, only half of the CLs under analysis consider it necessary to include an explicit rule on dictionary usage In the author's opinion, this is not an oversight Rather, this rule is understood implicitly in the other CLs
- SE rule 1 13 "Make your instructions as specific as possible" is shared by three other CLs
- SE rule 2 1 "*Do not make noun clusters of more than three nouns*" is shared by five other CLs Of the CLs that have a rule specifying the permissible size of noun clusters, two simply advise avoiding long noun clusters without specifying a number, another CL allows four nouns, while the remaining three allow three nouns

- SE rule 2 3 "When appropriate, use an article (the, a, an) or a demonstrative adjective (this, these) before a noun" is common to six other CLs
- Six CLs share a rule regarding the use of the gerund, or, more specifically, they recommend avoiding it AECMA SE does not have a specific rule which governs the avoidance of gerunds, but this is covered by rule 3 1 "*Only use those verb forms that are listed in the dictionary*" The gerund is not listed in the AECMA SE dictionary
- SE rule 3 6 "*Use the active voice*" is shared by six other CLs The rule takes the form of "*Do not use the passive*" (three CLs) to "*Avoid using the passive*" in the other three CLs
- Five CLs share a rule which recommends that relative pronouns such as "who", "which" or "that" should not be omitted AECMA SE does not have a corresponding rule

To summarise, the following linguistic phenomena are governed by rules in at least half of the analysed CLs dictionary usage, specificity of information in instructions, noun cluster size, article usage, gerund usage, passive voice and relative pronoun usage

1.4.6 Unique Rules

In the preceding section, rules that are common to multiple CLs are highlighted It is also interesting to examine the number of rules that are *unique* to each CL, i e rules which do not have a precise replica in any of the other CLs under analysis Table 1 3 highlights the proportion of rules that are unique to each CL

Controlled Language	***Proportion of Unique Rules***
AECMA SE	58%
ACE	83%
CL 1	51%
CL 2	48%
CL 3	30%
CL 4	32%
CL 5	42%
CL 6	50%

Table 1 3 Proportion of Rules Unique to Each CL

The two most noteworthy figures in Table 1 3 are the lowest and highest percentages of unique rules CL 3 has the lowest proportion of unique rules (30%), and CL 4 is not far off this figure with 32% The explanation for this fact is not immediately obvious However, it is known that CL 3 draws heavily on Simplified English rules with the effect that many of the non-unique rules in CL 3 are common to Simplified English The same cannot be said of CL

4: it is not derived from Simplified English and the rules it shares with other CLs are, in fact, shared mostly with CL 2, CL 3 and CL 5 and not with AECMA SE.

ACE has a significantly higher proportion of unique rules in comparison with the other CLs (83%). The explanation for this is that ACE sets itself apart from the other CLs in the analysis in terms of its objectives and this is reflected in the uniqueness of the rule set. As mentioned earlier, ACE (along with its successor PENG) is the only CL known to the author which focuses on "translating" a natural language CL into an artificial language. ACE's end-product, a software specification which is expected to be very specific and highly accurate, is quite dissimilar from other CLs where the end-product is usually a document for information purposes. The requirement for specificity and accuracy is reflected in many of the ACE rules, e.g. "*The principle of distribution in coordination states that if the complement of a (negated) verb consists of a coordination of phrases then the (negated) verb is distributed to each phrase. The following elements can be distributed: "is", "is not", finite full verb, "does not" + full verb, "does not". Non-finite elements (e.g. "not" on its own or adjectives alone) can not be distributed*'. ACE not only tells the author what constructions are allowed or disallowed (i.e. construction rules), e.g. "*Do not omit the relative pronoun "who", "which" or "that"*", but, as previously mentioned, it also includes principles of interpretation (i.e. interpretation rules). In doing so, it exercises an element of control which is not usually applied in other Controlled Languages.

1.4.7 Rule Analysis Conclusions

This analysis reveals that there is only one rule that is common to *all* CLs in the analysis, i.e. the rule which promotes short sentences. In addition, there are only seven rules that are common to 50% or more of the CLs. Despite the small number of CLs in this analysis, one would expect to find a greater number of common rules, in particular across the CLs classified as MOCLs. This suggests that the definition of CLs is largely individual and, more importantly, that the linguistic phenomena considered worth controlling vary from one organisation to the next.

1.5 CONCLUSIONS

Chapter 1 introduced the concept of Controlled Language, giving an historical overview and listing some well-known CLs. The justification for implementing CLs was discussed and this was offset against some of the known disadvantages. An analysis of eight CL rule sets for English was summarised in this Chapter. The aim of this analysis was to establish to what extent CL rules for English intersected. The results indicated a low degree of intersection, a result that has important implications for the current research. If the object of our investigation is to establish whether applying CL rules to an English source text (ST) leads to a reduction in the effort required to post-edit that text after machine translation, then we first need to apply some, ideally representative, CL rules to the ST As we have seen, however, it is not possible to identify a set of representative CL rules, and some compromise will be necessary This idea is discussed in detail in Chapter 2 where the concept of machine translatability and its relationship to Controlled Language is expanded upon

Chapter 2

2. CL, MT & TRANSLATABILITY

2.1 INTRODUCTION

This chapter examines the relationship between Controlled Language and Machine Translation. Different types of CL/MT users are discussed and we address the question of why Controlled Language is useful for Machine Translation by describing two approaches to reducing ambiguity using Controlled Language. Details are given on the types of lexical, grammatical and structural constraints imposed by CL rules in MT environments. The advantages of developing CL technology in conjunction with MT technology are explored and a number of different approaches to assessing "translatability" in a CL/MT framework are presented. Finally, the issues involved in post-editing in a CL/MT framework are explored at a preliminary level.

2.2 THE USERS

Groups interested in both Controlled Language *and* Machine Translation can be divided into three broad categories:

1. Users/Developers
2. Users
3. Developers

The first group includes companies who develop CL technology and use it in-house for research and/or production purposes. IBM is a good example of a user/developer (Bernth 1997, 1998a, 1998b, 1999a, 1999b). Academics involved in MT/CL research could also be placed in this first group. The second group includes companies who purchase CL technology and services from external sources for in-house deployment. Development of CL software is not generally undertaken by this second group, although they are involved in customising rules, integrating CL with in-house technology, for example Translation Memory and Machine Translation systems, and creating and maintaining lexica and rules. GM, when they were implementing the CASL project (Godden 2000), and Sun Microsystems (Wells-Akis 2002) are both examples of this type of user. The third group consists of companies who develop CL technology, and sometimes also MT technology, for commercial purposes. Such companies do not generally use the technology for production purposes, but are involved in providing consulting services and in customising rules and lexica. Smart Communications (Smart 1988), Acrolinx (http://www.acrolinx.com), Tedopres (http://www.tedopres.nl) and the IAI (Institut der Gesellschaft zur Förderung der

Angewandten Informationsforschung – Reuther 1998, 2003; Reuther and Schmidt-Wigger 2000; Schütz 2001) are included in this category.

2.3 WHY IS CONTROLLED LANGUAGE USEFUL FOR MACHINE TRANSLATION?

The ambiguous nature of language and the problems that ambiguity can cause for machine translation in general have been well-documented (Hutchins and Somers 1992; Arnold et al. 1994; Trujillo 1999). Rather than reproduce a general discussion on ambiguity and MT here, the focus will be on how CL scholars have tackled the problem of ambiguity in the context of machine translation. I have taken two examples of somewhat different approaches. The first example is KANT Controlled English and the KANT MT system and the second example is IBM's EEA and the LMT system. These systems were chosen because published descriptions of their methods for disambiguation were available.

2.3.1 Reducing Ambiguity using KANT Controlled English

Ambiguity resolution in the context of developing Kant Controlled English (KCE) for the KANT MT system is discussed in Baker et al. (1994), Mitamura and Nyberg (1995), and Nyberg and Mitamura (1996). The KANT system is classified as a Knowledge-Based MT system (KBMT), but the authors state that the principles of KCE are not restricted to KBMT (Mitamura and Nyberg 1995). In Nyberg and Mitamura (1996: 77), the types of ambiguity that must be dealt with, in general, are listed as

- Syntactic (multiple syntactic analyses)
- Lexical (multiple parts of speech, multiple meanings)
- Referential (WH-forms, WH-movement, pronouns, clitics etc.).

KCE deals with ambiguity on two of these three levels, i.e. lexical and syntactic (Mitamura and Nyberg 1995).[14] The constraints at each of these levels is explained in detail below.

2.3.1.1 LEXICAL CONSTRAINTS

According to Nyberg and Mitamura (1996: 78), the single most useful way to improve the accuracy of a knowledge-based MT system is to limit lexical ambiguity. Three strategies for limiting lexical ambiguity are recommended in Mitamura and Nyberg (1995). Firstly, meaning should be limited per word and part of speech (POS). Secondly, when a lexical item has more than one meaning in a domain, synonyms should be identified and used as

[14] Referential ambiguity is not dealt with in any detail by the authors because their work to date concentrates primarily on descriptive and instructive texts in technical domains where there is little or no need to support long-distance dependencies introduced by WH-words, pronouns and relative clauses (Nyberg and Mitamura 1996: 81).

alternatives Thirdly, terms whose ambiguity cannot be resolved with either of the previous two methods should be labelled for interactive disambiguation by the author

Additional lexical constraints are specified in order to reduce ambiguity The authors recommend that orthographical conventions are agreed with technical writers to avoid variation in spelling Rules governing the use of function words such as pronouns, determiners, reflexives, quantifiers and conjunctions, should be specified In particular, the use of conjunctions and pronouns should be restricted since they are a significant contributor to ambiguity Additionally, the use of participal forms, i e forms ending in *–ing* and *–ed*, should be restricted For example, instead of writing "when starting the engine" one would write "*when you start the engine*" Mitamura and Nyberg (ibid) also recommend that the *–ed* form should not be used to introduce a relative clause without explicit use of a relative pronoun So, instead of writing *"the pumps mounted to the pump drive"*, one would write *"the pumps that are mounted to the pump drive"*

An additional recommendation in Nyberg and Mitamura (1995) is to use a semantic domain model to limit parsing complexity during source analysis Semantic domain models are also known as semantic nets, which are defined by Trujillo (1999 7) as "collections of concepts linked together through a variety of relations " Arnold et al (1994 191) also comment on the use of a semantic domain model in the KANT system

> *Essentially, the premise is that high quality translation requires in-depth understanding of the text, and the development of the* domain model *would seem to be necessary to that sort of deep understanding*

The domain model specifies the relationships between all concepts in the domain This "knowledge" about relationships is used to help disambiguate the input for machine translation

The feasibility of using a semantic domain model is addressed by Nyberg and Mitamura (1996 82) They say that encoding a large number of semantic relations requires a cost-effective combination of automated acquisition from corpora, manual encoding and generalisation via semantic hierarchies If the domain is large and there is only a small amount of available text, then manual encoding might be prohibitive On the other hand, if a large corpus is available then corpus analysis techniques can be used to extract the domain model

Interestingly, Mitamura and Nyberg (1995) point out that the size of the vocabulary in KCE is not limited Only lexical and grammatical constructions which are unnecessarily complex are ruled out This, they argue, results in a language which is expressive enough to author technical documents, but which is limited in complexity

2.3.1.2 Grammatical Constraints

In Mitamura and Nyberg (ibid) two general types of grammatical constraints for the reduction of ambiguity are recommended the first places restrictions on the formation of complex phrases and the second on the structure of sentences

Phrase-Level Constraints

The authors make the following recommendations for phrase level constraints

- Verb + particles (e g verb (V) + preposition (P) or V + adverb (ADV)) should be avoided and replaced with single word verbs For example, the verb phrase *turn on* can be replaced with *start*
- Coordination of verb phrases should be made explicit For example, instead of writing *Extend and retract the cylinders* one would write *Extend the cylinders and retract the cylinders*
- Conjoined prepositional phrases (PPs) should be made explicit, e g *Five cubic metres of concrete and of sand* instead of *Five cubic metres of concrete and sand*
- The determiner should be made explicit in noun phrases (NPs)
- Nominal compounding should be avoided, unless it has been catered for in a rule in the MT system
- Quantifiers and partitives may not appear alone and must modify a nominal head, e g write *Repeat these steps until no bolts are left* instead of *Repeat these steps until none are left*

Sentence-Level Constraints

Mitamura and Nyberg (ibid) make the following recommendations regarding sentence-level constraints

- The two parts of a conjoined sentence should be of the same "type"
- Both clauses in complex sentences using subordinate conjunctions must contain a subject and a verb A subordinate clause should be able to stand on its own if the conjunction is removed
- Ellipsis should be ruled out in general However, elliptical phrases such as *if necessary* or *if equipped* may be required Mitamura and Nyberg recommend that such phrases should be defined as a "closed class" in the Controlled Language definition

- A relative clause should always be introduced by the pronoun *that* or *which*
- WH-questions should be ruled out wherever possible However, some domains require their use If this is the case, the authors recommend that they are re-phrased as direct questions using *do* or *be*
- Rules for consistent and unambiguous use of punctuation marks should be specified

2.3.1.3 Structural Constraints

Even if all known measures for reducing ambiguity are thoroughly implemented ambiguity may still occur, leading to incorrect MT output Mitamura and Nyberg discuss one additional mechanism for disambiguation, i e interactive disambiguation by the author using SGML (Standard Generalized Markup Language) tags

SGML tags provide structural information about the text and can be useful in resolving ambiguity If the MT system fails to disambiguate an input sentence, it is flagged and highlighted to the author who then tries to resolve the ambiguity with the help of the SGML tags The authors explain this method using the following example *Secure the gear with the twelve rivets* In this sentence the PP *with the twelve rivets* could modify either *secure* or *gear* An SGML tag can be inserted to indicate the right interpretation, i e *Secure the gear with* `<attach head="secure" modi="with">`*the twelve rivets*

2.3.1.4 Processing Constraints

In Nyberg and Mitamura (1996 81), an additional recommendation is made regarding processing constraints where the authors suggest controlling the complexity of the input sentence by setting a predetermined amount of time required for the analysis of each sentence If this time elapses and the analysis is incomplete, the system signals that the sentence is too complex and should be rewritten Another way of implementing this is by limiting the computer memory available for the analysis of each sentence The challenge, according to the authors, is to find the right threshold so that only overly complicated sentences are ruled out

2.3.2 Reducing Ambiguity using EEA

The treatment of ambiguity by IBM's EEA is primarily reported in Bernth (1998b) and also touched on in (Bernth 1997, 1999b) As already mentioned, the EEA checker is known as "a grammar checker++" (Bernth 1997 159) because it performs traditional grammar checks as well as CL checks The primary motivation for this is an unwillingness to impose too many restrictions on the writers So, most "standard English" constructions are allowed, with a few exceptions (Bernth 1999b) A second motive is the fact that while CL checkers can impose tight linguistic restrictions on writers, they sometimes allow ungrammatical content to pass through

> *Controlled Languages have been invented to solve the problems associated with readability and translatability, with slight regard to ensuring grammaticality*
>
> *(Bernth 1997 160)*

Therefore, performing standard grammar checks is beneficial, according to Bernth

As mentioned in the previous section, authors can be involved in disambiguation in some circumstances in the KANT system, but this involvement is limited to ambiguous constructions which the system has failed to resolve after several attempts The IBM approach is to include authors in the disambiguation process as much as possible by flagging ambiguities and asking them to disambiguate interactively The EEA Checker, therefore, only checks for some instances of structural ambiguity, complexity and vocabulary violations In total, it performs 40 checks (Bernth 1997)

2.3.2.1 LEXICAL CONSTRAINTS

EEA has a general dictionary of 80,000 words The user can specify which dictionary he or she wants to use during the checking stage Checks are carried out on restricted words, acronyms and abbreviations, and controlled vocabulary In addition to these three, checks for misspellings, unknown words and duplicate words are also performed

2.3.2.2 GRAMMATICAL CONSTRAINTS

The syntactic checks performed by the EEA Checker include checks for a lack of parallelism in coordination and list elements, ambiguous coordination, ambiguous attachment of non-finite clauses, subject ellipsis, passives, double negatives, long sentences, wrong pronoun case and long noun strings (Bernth 1997, 1998b)

Examples of how EEA treats ambiguous coordination are provided in Bernth (1998b) Having written the sentence *Give the data or information sheet to your manager*, the user is asked to disambiguate between the following two interpretations

1 the information sheet or the data OR

2 the information sheet or the data sheet

Depending on the user's choice, the output from the LMT system (in German) can be either

1 Geben Sie Ihrem Manager das Informationsblatt oder die Daten OR

2 Geben Sie Ihrem Manager das Informationsblatt oder Datenblatt

Similarly, examples are given for disambiguating the attachment of non-finite clauses Having written *A note is forwarded to the user requesting the correct information*, the user is asked to disambiguate between

1 A note that requests the correct information is forwarded to the user OR

2 A note is forwarded to the user that requests the correct information

Again, depending on the user's choice, the output can be either

1 Eine Notiz, die die richtige Information fordert, wird an den Benutzer weitergeleitet OR

2 Eine Notiz wird an den Benutzer weitergeleitet, der die richtige Information fordert

2.3.2.3 Structural Constraints

EEA checks for punctuation errors such as missing commas and hyphens Bernth (1997) echoes Mitamura and Nyberg (1995) by recommending the use of SGML tags in the disambiguation process These are used by the EEA Checker to help with sentence segmentation and with the identification of tables, displays and revised text

2.3.2.4 Summary

In this section I have outlined two, quite different, approaches to the resolution of ambiguity by CLs Both the IBM and Carnegie Mellon (KANT) approaches tackle ambiguities on the lexical, syntactic and structural level However, IBM takes the view that the traditional definition of CL is highly restrictive and yet does not cater for ungrammatical input Their requirement is for some loose CL checking combined with some grammatical checking The KANT system, on the other hand, imposes a greater number of highly restrictive rules Both approaches allow for author involvement in the disambiguation process but differ in the extent to which the author is involved, with the KANT system making use of author disambiguation after the main processing stage and the IBM system involving the author during the main processing stage

2.4 Development Considerations

As already mentioned, developers of CL technology can sometimes also be involved in the development of MT technology and, in such circumstances, advantage can be gained by re-using components of the MT system for the CL system as reported in Adriaens (1994) and Knops and Depoortere (1998)

Adriaens (ibid) describes the development of a CL within the SECC project ("Simplified English Grammar and Style Checker/Corrector") The main objective of this project was "the development of a tool for technical writers who produce documents in a variant of Simplified

English" (Adriaens 1994 78) Although this suggests that the focus of the project was to provide supporting technology for monolingual authoring, machine translation, in the form of the METAL system, also had a role to play

The SECC team viewed Controlled Language as an MT problem Their aim was to produce a new language pair, English-Simplified English, where SE would then act as an interlingua for transfer into other target languages To create the new language pair, they developed a new transfer module as they would have done for any other new source-target language pair METAL used a small part of the existing English generation component for generating SE, but since SE was a subset of English, its generation component was not as extensive as for other language pairs However, the SECC team encountered some problems when re-using METAL's classical analysis-transfer-synthesis approach for SE METAL would overwrite the input string with the output string and this meant that error diagnosis was problematic

Knops and Depoortere (1998) also discuss the advantages of re-using MT components for CL development Their system, LANTMARK, was in fact based on SECC and their core MT engine was also METAL Knops and Depoortere (ibid 43) list the benefits of designing a CL checker in an MT framework as

- Re-use of existing MT components
- Re-use of converters for separating text from formatting
- Re-use of software for extracting and re-inserting translation units
- Re-use of the Client/Server architecture for batch processing
- Re-use of the English analysis component and lexicon [15]

The authors observed, however, that further development of the CL application was not possible without changing some aspects of the MT system For instance, the system had to be able to process ungrammatical input and to make use of text structure information (using SGML tags), developers had to find a way of preserving the link between the input and output sentences, and, finally, they had to introduce interactive checking to complement batch processing

In addition to the changes listed above, a "CL switch" was implemented which triggered one of two operating modes, i e normal mode and Controlled English mode In Controlled English mode, the grammar excluded certain non-conformant interpretations Also, filtering during morphological analysis meant that only words contained in the controlled

[15] Adriaens goes into some detail on the *difficulties* involved in re using the general English lexicon for Controlled Language (1994 81 82)

lexicon were allowed and all other interpretations could be ruled out The expectation was that this would improve the output and processing speed

> *Translation quality will profit from CL biased disambiguation, while the exclusion of non-conformant structures reduces the search space for analysis, and thus enhances the translation speed*
>
> *(Knops and Depoortere 1998 46)*

In this section on development considerations, I have represented two opinions on developing CL in an MT environment Both Adriaens (1994) and Knops and Depoortere (1998) enunciate the benefits to be gained from combining CL development with MT development the analysis component can be reused, which leads to consistency in results, the CL can be viewed as an interlingua for MT architectures that allow this, parts of the lexicon and other components which deal with processing and formatting can be reused However, problems are also associated with this development approach For example, re-use and maintenance of the lexicon is not as straight-forward as it might seem and the MT system itself might require some customisation to allow, for example, the output to be viewed in the manner required Overall it seems that a combined development environment is both sensible and advantageous

2.5 Confidence Indices

Within the domain of CL and MT, some researchers have examined the possibility of calculating a “confidence index” for MT output (Gdaniec 1994, Bernth 1999a, 1999b, 2000, Underwood and Jongejan 2001) The MT confidence index is assigned to the output by a human translator/post-editor who rates the output according to prescribed criteria While this type of evaluation is beneficial, greater benefit can be gained from assigning a confidence index to the source as opposed to the target text Such a method boasts the following advantages

- Problematic source segments can be identified and pre-edited before MT with the result that a higher number of segments are translated to a satisfactory quality level
- If the source text is to be translated into multiple target languages, fixing one error in the source means improving the output and reducing the post-editing required for multiple target languages
- CL checking technology can be re-used to create a “translation confidence index” (TCI)

- TCI values can be calibrated to take account of specific source and target language pairs, making the TCI value more accurate and more useful
- TCI values can also be calibrated to take account of specific MT systems, again making the TCI value more accurate and more useful

The next section reports on the research findings published to date concerning the topic of translatability and translation confidence indices and elaborates on some of the advantages listed above

2.6 TRANSLATABILITY

Several authors list "translatability" as one of the main goals of CL (Wojcik and Hoard 1996, Reuther 1998, Means and Godden 1996) However, no definitions are given for the concept of "translatability" [16] At the time of writing, five authors have been identified who have written in detail on translatability assessment and whose work appears in the following sources Gdaniec (1994), Bernth (1999a, 1999b), Bernth & McCord (2000), Underwood and Jongejan (2001), and Bernth and Gdaniec (2001) A summary of the discussion in these papers is presented below

2.6.1 Gdaniec on Translatability

In Gdaniec (1994) the notion of a "translatability index" (TI) is introduced The TI is based on the gross statistical properties of a document, rather than on the low-level parsing of sentences It is important to note that the author sees the TI as having "relative significance" and not "absolute significance" In other words, the TI can tell if one text is more suitable for MT than another, but it cannot tell if a document will produce acceptable MT output

The calculation of the TI is achieved by identifying so-called "negative" sentence properties and assigning penalties The software program which calculates the TI starts off with a value of 7, i e the highest possible score, and then subtracts values as each negative property is identified Gdaniec reports that 39 sentence properties are used in the scoring of the TI, some of which are source-language specific while others are common to source and target languages Some of the negative sentence properties included in the TI calculation were derived from knowledge of the MT system the author was working with, i e Logos Where it was known that certain grammatical properties were not handled well by Logos, these properties were added to the TI calculation program This means that the TI calculation

[16] Means and Godden do however state that 'translatability" is a factor of the "readability factor" (1996 107), as mentioned in Chapter 1

was not only source and target-language specific, but also specific to the MT system The negative sentence properties identified by Gdaniec are

- Words not contained in the dictionary
- Short parentheses
- Coordination
- Homographs
- Interrogatives
- Unmatched parentheses
- Dependent and relative clauses
- Complement sentences
- Noun and verb form ambiguities (for German as source)
- Nested participle constructions modifying nouns (also for German as source)
- "Suspicious" or difficult pronouns

And, specifically, for German as a SL

- Certain ambiguous words (“also”, “desto”, “wie” and “da”-compounds)
- Sentences beginning with “dass”
- Words that reflect syntactic ambiguities, such as participles
- Words that can be both pronoun and determiner
- Inverted sentences functioning as conditionals
- Certain pronouns and possessive determiners (“sie”, “ihm”, “er”, “ihr”, “sein” etc)
- Strings of noun-compounded nouns

A different method is used to calculate the TI for English and German source documents For English source documents, the “text TI” is the *average* of all sentence TI’s However, for German source documents, the “text TI” is the average of the sentence TI’s *plus* “a special document TI” which assigns additional penalties for certain criteria which characterise the text as difficult for the Logos system The author does not explain why an additional penalty is required for texts with German as a source language The answer to this could be that the Logos analysis component for German is perhaps not as robust as the analysis component for English with the consequence that complex German grammatical structures are not transferred correctly into the target language? It is also possible that the author did not get to a stage in her research where an equivalent document TI for English was devised

Gdaniec maintains that there is a correlation between the gross properties of a text, e g sentence length, degree of syntactic complexity and discourse characteristics, and the quality of MT output She devises a scale for measuring the quality of MT output called the "QI" or "Quality Index" The QI rates grammatical correctness, understandability and preservation of information, but not style Like the TI, the QI is based on a scale of 1-7, where the numbers indicate the following (Gdaniec 1994 105)

- **7** - No corrections required
- **6** - One or two minor changes are required No reference to the source is required because the sentence is understandable despite the errors
- **5** - Several minor changes are required The solution is relatively obvious, but reference to the source may be desired for confirmation
- **4** - There are multiple solutions, or the solution is not obvious from reading the translation Reference to the source is required to correct the sentence
- **3** - Major changes are required Reference to the source is definitely required to make the changes
- **2** - Large parts of the sentence must be retranslated
- **1** - The entire sentence must be retranslated, with virtually nothing salvageable from the raw translation

According to Gdaniec, texts with a QI value of 5 or higher are widely accepted as input for post-editing

Gdaniec reports on tests carried out using 541 English sentences (with German and French as target languages) and circa 1,300 German sentences (with English and French as target languages) The sentences were taken from different text types and subject matter areas A relatively high correlation between TI and QI scores was found approximately 94-95%, with a few exceptions Even with the exceptions, the documents with the lowest QI score also had the lowest TI score

Attention is drawn to the fact that the correlation between the TI and QI will vary for different language pairs because some target languages have a higher "free ride" factor (in the words of Gdaniec) than others For example, if the boundaries of an English relative clause are determined incorrectly, this may not affect a French target, but will produce an incorrect German target. This means that the TI has to be adjusted to reflect the target language.

Gdaniec acknowledges the subjectivity involved in assigning translatability and quality indices:

> *In designing this project, we took into account the fact that measures of translatability and the evaluation of translations are necessarily subjective and therefore, while measures are clearly needed and may indeed be useful, they are impossible to fix in any absolute sense.*
>
> *(Gdaniec 1994: 98)*

Nevertheless, she is confident that the value of such an exercise will become clear over time:

> *The long-range feasibility of a TI hinges on the establishment of an index that takes on significance to the users over time, as they work with it and find it useful. A TI is useful if it can provide the user with a measure that correlates reasonably well with the quality of the MT output over a period of time. In particular, the index will establish its usefulness if, as the users manipulate the source document to improve the index, they experience a corresponding improvement in the quality of the MT output.*
>
> *(Gdaniec ibid: 99)*

In the context of manipulating the source text to improve MT output, Gdaniec suggests that the TI program could advise the user on how to increase the value of the TI by providing feedback like, for example, "On average, the sentences are too long" or "There are too many "ja" and "gerade"-type words" etc. In effect, this would turn the TI application into a type of Controlled Language checker.

2.6.2 Bernth and McCord on Translatability

Bernth and McCord contribute to the literature on translatability by describing IBM's efforts to develop and tune a "Translation Confidence Index" (TCI) (Bernth 1999a, 1999b; Bernth & McCord 2000). The TCI is described as "...a function that assigns to each source language segment a number that estimates the confidence that the MT system can translate that segment well" (Bernth and McCord 2000: 89). Professional translators use the values assigned by the TCI to filter out unwanted MT output. In theory, the TCI value could also be used as a quality indicator by those who do not know the source language, for example, internet users who machine translate web pages.

In general, the TCI is based on a number of factors including language pair and language distance.[17] In Bernth and McCord (ibid), a very detailed account is given of the linguistic phenomena used for calculating the TCI and the values of penalties assigned for each phenomenon, which I will summarise here.

[17] Bernth (1999b) mentions Odlin (1989) here, saying that although Odlin's description of the impact of language distance is based on the use of language by humans, it is even more true of machine translation.

The TCI value lies between 0 and 10, where 0 implies no confidence in the translation and 10 implies complete confidence The computation of the TCI value is regarded by Bernth and McCord as "an attempt to approximate the way an expert human translator would rate the MT system's translation on a scale of 0 to 10" (ibid 89) In fact, the idea of calculating a TCI is closely related to the idea of automatic MT evaluation This work differs from previous work on evaluation because the evaluation is carried out *during* the translation process and, consequently, the translation for a given segment can be abandoned if the value goes below a certain threshold (ibid 90) The authors emphasise that heuristics are used in the calculation of the TCI value with the result that a value of 10 does not guarantee a perfect translation Another important consideration is that the TCI value is a measurement for a *particular* MT system (LMT in this case) However, it is claimed that the authors' specific method of calculating a TCI value may apply across different MT systems

TCI values are assigned as follows each segment starts with a value of 10 and penalties, which are floating point numbers, are assigned depending on the linguistic phenomena encountered in each segment and the seriousness of the problem introduced for machine translation by those linguistic phenomena The impact of each problem varies from language pair to language pair For this reason, exact penalties are set in a language-pair-specific profile This profile contains two sets of specific data (1) the penalties that are available with the TCI and (2) the *adjustment coefficients* available for each penalty The TCI engine computes a number associated with the penalty and this value is multiplied by the adjustment coefficient before the penalty is applied to the overall score

Penalties may be applied during any part of the translation process source analysis, lexical transfer, structural transfer and morphological generation The most serious problem with the lexical transfer phase is when there is no entry in the lexicon for a specific word Bernth (1999a 124) points out that the more complex the entry, the more likely it is to be recorded in the lexicon, but also the more likely it is that an error occurred when that entry was made in the lexicon Informal studies by Bernth revealed that this is a factor which should not be discounted, so the number of transfer elements is taken into account in the TCI calculation

Bernth discounts the problems caused by target morphological generation as "very insignificant" (ibid 124) because target morphology is well-defined and limited Problems that arise in this area are usually caused by previous steps While highly inflected parts of speech can be a problem, this is catered for by the TCI which assigns a higher penalty when such a POS occurs

Source analysis plays the most important role in TCI computation because it is the most non-deterministic and most error prone part of the MT process and also because errors made at this stage tend to be carried over into other stages (Bernth and McCord 2000 90-92) The main steps in source analysis where problems can occur are identified as segmentation and tokenisation, lexical and morphological analysis and syntactic analysis Table 2 1 is adapted from Bernth and McCord (2000) It illustrates what linguistic phenomena are counted in each of these source analysis steps (the "Problem Types" column) and shows the formulae for penalty calculation

Problem Type	**Penalty**
Lack of Initial Capitalisation of Segment	1
Abbreviations	0 1
Punctuation	0 5
Footnotes	1 5
Segment Length	Penalty applied is a function of the segment length
Lexical Analyses L	If L = 0, then 0 3, else 0 01 * L * *s*[18]
Parts of Speech	*number_of_parts_of_speech* * *s*
Noun-Verb	0 07 * *s*
Determiner/ Pronoun-Noun/Verb	0 07 * *s*
Infinitive/Imperative Verb-Noun	0 1 * *s*
Adjective Noun-Noun	0 5 * *s*
Infinitive/Imperative Verb/Adjective-Noun	0 1 * *s*
Proper Noun-Noun	0 1 * *s*
"To"-Infinitive Verb/Noun (segment length > 3)	1 5
Coordinating Conjunction	0 5
Problematic Words	0 1
Failed Parse	5
Parse with Unfilled Obligatory Slots	1 0
Many Parses	Penalty is applied according to the number of parses generated by the MT system
Identical Parses with Different Word Senses	0 5
Close Parses	0 25 * *s*
Parsescore	(0 5 * *parsescore*) / *segment_length*
Missing Subject	0 1
Missing Hyphen	0 1
Lack of Subject-Verb Agreement	0 1

[18] s = *shortfactor* which is an additional penalty for segments of four words or fewer where there is more than one lexical analysis for one or more of these words See Bernth and McCord 2000 for more details

Problem Type	Penalty
Wrong Comparative/Superlative	0 1
Long Noun Groups	0 1
Missing "that"-Complementizer	0 05
Passive Construction	1 5
Non-finite Verb	0 5
Potentially Wrong Modification in Subjectless VP	0 2
String of Prepositional Phrases	0 05
Double Ambiguous Passives	0 1
"For-to"-Constructions	0 5
Time Reference	0 2
Prepositions with Objects	3
Lexical Time Usage	(0 01 * *time) / segment_length*)
Syntactic Time Usage	(0 002 * *time) / segment_length*)
Pointer Space Usage	(0 00001 * *time) / segment_length*)
Character Space Usage	(0 001 * *time) / segment_length*)

Table 2 1 IBM TCI Penalties

Bernth and McCord detail the type of testing they have done with the TCI penalties Their aim was to obtain a profile with adjustment coefficients determined from regression analysis (2000 96) To accomplish this, they used two sets of training sentences, with 200 sentences in each set The first set was composed of online newspaper articles and the second of web page news stories The principal idea behind the regression training is to tune the TCI so that it is close to the scores that would be assigned by a human evaluator of MT They then compare the results for a regression-trained TCI to a hand-tuned TCI A threshold TCI value of 7 is set Any segments with a TCI value lower than 7 are not shown to the translator The main conclusions from this comparison are that

- the hand-tuned TCI is more successful at not showing bad results to the user (88 2% success rate vs 71 4% for regression-trained TCI) but this is at the expense of not showing as many results in total to the user (only 28 3% for hand-tuned TCI vs 58 3% for regression-trained TCI), i e precision is higher for the hand-tuned TCI, but recall is lower
- Taking all statistics into account, Bernth and McCord judge that the regression-trained TCI performs better than the hand-tuned TCI
- The authors state that they intend to follow the regression-training path to further refine the TCI algorithm

2.6.3 Underwood and Jongejan on Translatability

Research on the topic of translatability is justified by Underwood and Jongejan (2001) by pointing out that MT output can sometimes be of such poor quality that it would be faster to produce the target text using a human translator than to post-edit the MT output Thus, it would be beneficial to be able to identify in advance those texts which are not suitable for MT

Similar to Gdaniec (1994), Underwood and Jongejan assess translatability on the basis of negative sentence properties, termed "translatability indicators"

> *The notion of translatability is based on so-called "translatability indicators" where the occurrence of such an indicator in the text is considered to have a negative effect on the quality of machine translation The fewer translatability indicators, the better suited the text is to translation using MT*
>
> *(Underwood and Jongejan 2001 363)*

Also similar to Gdaniec's approach is the fact that translatability is assessed by a shallow and rapid analysis of the source text, rather than a full low-level sentence parse, which consequently leads to a trade-off between speed and robustness, on the one side, and accuracy on the other It seems that whereas the output from Gdaniec's translatability assessment software is a TI value alone, Underwood and Jongejan's tool also produces an annotated text This is of benefit to the user who can then use the annotated text to ascertain how the TI tool performed its analysis

Underwood and Jongejan (ibid) build on the TI indicators listed by Gdaniec, adding

- Structural ambiguity caused by prepositional phrase (PP) attachment, relative and other sub-clause attachment and multiple coordination
- Compounds comprising three or more nouns
- Sentences without finite verbs
- Lexical ambiguity
- Sentence length (both very long and very short)

As with Gdaniec, the authors work with a specific MT tool, *PaTrans*, which is primarily used for translating patents for the language pair Danish-English Consequently, specific indicators are also identified which cause problems for the PaTrans MT system, i e sentence-initial prepositional phrases, adverbs, sub-clauses

Underwood and Jongejan use the Brill part-of-speech (POS) tagger[19] in conjunction with the Penn Tree Bank tag set[20] to identify the following general phenomena

- No verb present
- No finite verb present
- Multiple coordination
- Long sentence (i e >25 words)
- Short sentence (i e <3 words)
- One or more nominal compounds (>2 nouns)

The analysis also tries to identify the following specific indicators

- Sentence over 25 words with at least one adverb
- Sentence-initial adverbs or subclauses
- Sentence-initial PPs and/or subclauses
- Non-sentence-initial PPs and/or subclauses
- PP headed by "of"

Their word-based analysis searches for

- Noun-verb homographs
- Adjective-verb homographs
- Adjective-noun-verb homographs

Pattern matching rules are then used to identify translatability indicators These rules are “simple” in that they rely solely on identifying the presence or absence of a particular tag and on calculating the length of a sentence For example, if the verb tag is not present in a sentence, then it is a verbless sentence

The formula suggested for calculating the TI value is

$$\int I_{ik} = \frac{m_{ik}}{1 + m_{ik}}$$

Underwood and Jongejan (ibid 365) explain this formula as follows if there are m_{ik} occurrences of an indicator i in sentence k, then the fractional indicator value I_{ik} is computed as shown above Thus, the occurrence of 6 conjunctions would give a TI value of 0 857, i e

[19] See http //www cs jhu edu/~brill/ for more information [last accessed January 25 2006]

[20] See http //www cis upenn edu/~treebank/ for more information [last accessed January 25 2006]

$$\int I_{ik} = \frac{6}{7}$$

The index value for translatability indicators present in a sentence will normally be a number between 0 and 1, although those indicators which can only occur once in a sentence, e.g. "no verb", "long sentence", will have a value of 1. In comparison, Gdaniec's TI values range from one to seven. However, it is not possible to compare the calculation methods as Gdaniec does not go into detail on how her TI values are assigned.

Once TI values have been calculated by Underwood and Jongejan, a process of weighting then occurs. Each TI indicator has a weight assigned to it which is dependent on the relative effect of that indicator on MT output. The authors give weights a value of between 1 and 100, although no explanation is given as to the reason for this. If a weight is assigned a value of *0,* then it is considered irrelevant and the pattern matching rule for the indicator is disabled. The TI of any sentence is computed by adding the weighted values for every TI indicator in that sentence. The *text* TI is then the average of all the sentences' TIs. The TI value then lies between 1 and 100, where 100 means that no TIs are present that would have an adverse effect on the MT output.

It is interesting to note that the authors suggest maintaining a different weight file for each MT system. Such weight files would reflect the weaknesses in analysis and transfer for each MT system. Presumably, it would also be beneficial to have a different weight file for each language pair for each MT system. In fact, it may also be beneficial to reflect the capabilities of the MT system regarding different text types in the weight file.

The TI assessment tool was tested by the authors using *Recall* and *Precision* as criteria. The results appeared positive. However, this may be attributed to the fine-tuning of the tokenisation and segmentation algorithms to suit each specific text-type, according to the authors. At the time of writing, no comparison between the TI values, the quality of MT output, and the post-editing effort had been done, but this was highlighted by the authors as potential future research: "...to fully evaluate the checker, it is necessary to [...] have data on the actual post-editing effort required to transform the raw MT output into publishable quality" (ibid: 367).

2.6.4 Bernth and Gdaniec on "MTranslatability"

Bernth and Gdaniec teamed up in 2001 to discuss "MTranslatability" or *Machine Translatability.* Their 2001 paper is a systematic analysis of the factors that can cause problems for MT. While the discussion of translatability in the context of CL is often bound to specific MT engines, Bernth and Gdaniec claim that their approach can be viewed as being more general Although all examples use English as the source language, it is their contention that the general principles can be carried over to other languages

The authors report on an analysis they carried out of several grammar and style checkers for English and German as source languages They categorised the checks into (a) Useful for MTranslatability, (b) not useful for MTranslatability, and (c) more or less harmful for MTranslatability They found that most of the checks fell into category (a) - Useful for MTranslatability However, they also found that some recommendations fell into category (c), e g recommendations on sentence variety Their conclusion was that grammar and style checkers demonstrate a limited usefulness in the preparation of a document for MT

Bernth and Gdaniec highlight the difference between readability and translatability Although authors sometimes assume that these two are synonymous, an experiment carried out by the authors shows that this is not the case The experiment consisted of editing some sample problematic sentences according to their MTranslatability criteria and analysing them for MTranslatability and readability The results showed improved clarity and translatability, but reduced readability scores This finding is in contrast with the results of a similar comparison conducted by Reuther (2003), which was already mentioned in Chapter 1

Bernth and Gdaniec describe automatic MTranslatability scoring in terms of the Logos Translatability Index and the IBM Translation Confidence Index Underwood and Jongejan's work on translatability is also mentioned but no details were available to the authors at the time of publication The authors contend that the parts of the IBM TCI calculation process that take source analysis into account give a picture of "the general MTranslatability" of a text

> *Hence, turning all non-source-language-specific factors off in the user profile in effect gives a MTranslatability score that can be independent of the target language*
>
> *(ibid 38)*

The MTranslatability rules put forward by Bernth and Gdaniec are available in Appendix A

2.6.5 A Comparison of Approaches

The four approaches to translatability computation reported above are similar in the following ways

- The target language is considered to be significant in calculating the translatability of a document
- The specific MT system being used is significant in calculating the translatability value

- Each applies weights to the penalties based on criteria such as TL or complexity of the linguistic problem.

The most significant differences between the four approaches can be summarised as:

- Bernth and McCord and Bernth and Gdaniec employ a low-level, segment-based method for TCI calculation, whereas Gdaniec calculates TCI values based on the text rather than the segment, and Underwood and Jongejan use a combined approach.
- This difference in approach leads to a more detailed penalty formula for Bernth and McCord (see Table 2.1 compared with the single formula presented by Underwood and Jongejan).
- The numeric ranges for the TCIs differ, making comparisons impossible. In the combined approach of Bernth and Gdaniec, no specific values for penalties are suggested.
- Bernth and Gdaniec take a more generic approach to translatability, claiming that their MTranslatability rules could apply to multiple source and target languages and multiple MT engines.

While it is useful to make a comparison of approaches based on generic features, the most significant comparison for our purposes looks at the linguistic phenomena identified as being important for the TCI calculation by all authors. In order to make such a comparison, the table of features used by Bernth and McCord will be used as a template (as this is the most detailed) and a comparison will be made with the linguistic phenomena listed by Gdaniec and Underwood and Jongejan.[21] It should be noted, however, that since the authors use different naming conventions to describe similar phenomena, there is a difficulty in providing an entirely accurate comparison. This problem is exacerbated by the fact that the approaches to translatability calculation differ, with one author (Gdaniec) preferring a more generic text-based approach, which is reflected in her use of terminology for categorising the linguistic phenomena considered by the TCI algorithm. In addition to this, it is important to note that Bernth and McCord are approaching translatability from a CL culture, whereas Gdaniec and Underwood and Jongejan are approaching it more from a generic MT standpoint.

I have tried to find correspondences, in so far as possible, between the naming conventions used by each author. I have provided their exact wording in italics and

[21] Bernth and Gdaniec (2001) has been excluded from this comparison because they do not discuss the calculation of a specific translatability value, but rather provide more generic rules on translatability.

parentheses so that the reader can make his/her own mind up as to the degree of accuracy in the comparison. The features below the thick black line in Table 2.2 represent the linguistic phenomena mentioned either by Gdaniec or Underwood and Jongejan which are not explicitly listed by Bernth and McCord. Finally, it is important to draw attention to the fact that just because there is an "×" beside a linguistic feature, this does not prove conclusively that the author does not take this phenomenon into account during calculation of the TCI. It simply indicates that this phenomenon is not explicitly mentioned in the papers referenced.

Problem Type	***Bernth & McCord***	***Gdaniec***	***Underwood & Jongejan***
Lack of Initial Capitalisation of Segment	✓	×	×
Abbreviations	✓	×	×
Punctuation	✓	×	×
Footnotes	✓	×	×
Segment Length	✓	×	✓ *(sentence length both very long and very short)*
Lexical Analyses	✓	✓ *(words not contained in the dictionary)*	✓ *(lexical ambiguity)*
Parts of Speech	✓	×	×
Noun-Verb	✓	×	✓
Determiner/ Pronoun-Noun/Verb	✓	×	×
Infinitive/Imperative Verb-Noun	✓	×	×
Adjective Noun-Noun	✓	×	×
Infinitive/Imperative Verb/Adjective-Noun	✓	×	✓ *(Adj-Noun-Verb homographs)*
Proper Noun-Noun	✓	×	×
"To"-Infinitive Verb/Noun (segment length > 3)	✓	×	×
Coordinating Conjunction	✓	✓ *(coordination)*	✓ *(multiple coordination)*
Problematic Words	✓	✓ *(words that reflect syntactic ambiguity such as participles)*	×
Failed Parse	✓	×	×
Parse with Unfilled Obligatory Slots	✓	×	×
Many Parses	✓	×	×
Identical Parses with Different Word Senses	✓	×	×

Problem Type	**Bernth & McCord**	**Gdaniec**	**Underwood & Jongejan**
Close Parses	✓	×	×
Parsescore	✓	×	×
Missing Subject	✓	×	×
Missing Hyphen	✓	×	×
Lack of Subject-Verb Agreement	✓	×	×
Wrong Comparative/Superlative	✓	×	×
Long Noun Groups	✓	✓ *(strings of noun-compounded Nouns)*	✓ *(compounds comprising 3 or more nouns)*
Missing "that"-Complementizer	✓	✓ *(complement sentences)*	✓ *(sub-clauses)*
Passive Construction	✓	×	×
Non-finite Verb	✓	×	×
Potentially Wrong Modification in Subjectless VP	✓	×	×
String of Prepositional Phrases	✓	✓ *(structural ambiguity caused by PP attachment)*	✓ *(sentence-initial PP)*
Double Ambiguous Passives	✓	×	×
"For-to"-Constructions	✓	×	×
Time Reference	✓	×	×
Prepositions with Objects	✓	×	×
Lexical Time Usage	✓	×	×
Syntactic Time Usage	✓	×	×
Pointer Space Usage	✓	×	×
Character Space Usage	✓	×	×
Ellipsis of finite verb	×	✓	✓ *(No finite verb present)*
Interrogatives	×	✓	×
Inverted sentence functioning as a conditional	×	✓	×
Dependent & relative clauses	×	✓	✓ *(Structural ambiguity caused by…relative and other sub-clause attachment)*
Adverbs	×	×	✓
Sentence over 25 words with at least one adverb	×	×	✓
Sentence-initial adverbs or	×	×	✓

Problem Type	***Bernth & McCord***	***Gdaniec***	***Underwood & Jongejan***
subclauses			
Sentence-initial PPs and/or subclauses	×	×	✓
Non-Sentence-initial PPs and/or subclauses	×	×	✓
PP headed by "of"	×	×	✓
Adj-Verb homographs	×	✓ *(Homographs)*	✓

Table 2 2 Comparison of Linguistic Features Used in TCI Calculation

As can be seen from Table 2 2, Bernth and McCord take many more linguistic factors into consideration in their TCI computation than the other two approaches The following are included in all approaches lexical analysis, coordination, long noun groups, use of subclauses and prepositional phrases

2.6.6 Other Contributions on Translatability

Kohl (1999) discusses improving readability and translatability by using "syntactic cues" which are defined as "elements of aspects of language that help readers correctly analyze sentence structure and/or to identify parts of speech" (ibid 149) He classifies definite and indefinite articles as significant syntactic cues in English and in many Western European languages The benefit of this approach over Controlled Language is expounded as follows (ibid 150)

> *In contrast to Controlled English, the syntactic cues approach does not impose inordinate restrictions on vocabulary nor on the range of grammatical constructions that are permitted And, when used with discretion, it doesn't result in language that sounds unnatural to native speakers*

Kohl offers a ten-step approach to implementing the Syntactic Cues strategy for increasing translatability, i e

- Make sure you do not use a telegraphic style
- Consider expanding past-participles using *that*
- Look for present participles and make changes [22]
- Search for *and* and consider whether you should insert an infinitive marker [23]
- Search for occurrences of *or* [24]

[22] Here he offers seven examples of where a change might be implemented For example if the ING word follows a verb such as *begin start* or *continue* that can take an infinitive complement then consider changing the –ING word to an infinitive

[23] *Six specific guidelines are offered here For example, if the and joins* two noun phrases, and if an adjective precedes the first noun, then consider whether that adjective modifies (a) both nouns or (b) only the first one

[24] Five guidelines are then offered on strategies that can be implemented for *or*, e g if *or* joins two verb phrases, consider inserting a pronoun or noun subject so that you will have a compound sentence or two separate sentences instead

- Search for long noun phrases and hyphenate compound adjectives
- Look for specific verb forms like *assume, ensure, indicate* and ask yourself if you could put the word *that* after these verbs to make the sentence structure clearer
- Look for the verb forms *give* and *assign* Consider whether there is an indirect object that could be made grammatically explicit by using the word *to*
- Search for *if* If the first *if*-clause is followed by a second conditional clause, then the second one should generally also start with *if*
- Look for adjectives following nouns and consider expanding the adjective into a relative clause

Kohl's steps, in particular one to six, echo the linguistic features identified as being problematic for translatability by Bernth, McCord, Gdaniec, Underwood and Jongejan However, he specifically distances his approach from Controlled Language (as indicated in the quotation above) and he does not approach translatability from an MT perspective For these reasons it was deemed inappropriate to include the syntactic cues approach in this research

Spyridakis et al (1997) provide details of an experiment carried out to test the translatability of documents written in Simplified English (SE) and non-SE The STs were translated into Chinese, Spanish and Japanese respectively Translations were then rated by native speakers of each language who were suitably qualified (according to the authors) and who demonstrated good communication skills in English (ibid 6) The translatability ratings included five measures

- Accuracy of the translation (1 was best, 5 was worst)
- Style match with the original document (1 was best, 5 was worst)
- Ease of comprehension (1 was best, 5 was worst)
- Number of major and minor errors
- Number of major and minor omissions

They found that "subjects who translated SE documents produced higher quality translations than those who translated non-SE documents" (ibid 7) However, this finding did not apply to Chinese where "there were no significant differences between SE and non-SE translations" (ibid 8) Spyridakis et al speculate that the reason for the lack of a notable improvement for Chinese might be due to the extensive changes in linguistic structure that a translation from English into Chinese would undergo These changes might effectively nullify the differences between an SE and non-SE source text However, they do not explain why this would not also apply to Japanese as a TL

Wells-Akis and Sisson (2002) discuss translatability from a business viewpoint They describe the successful implementation of CL and MT at Sun Microsystems between 1999 and 2002, during which "translatability guidelines" were followed by one of their technical publications groups No formal definition of what they mean by translatability is offered Thirty guidelines were implemented (these are not described in detail) and they found the rule that limits sentence length to fewer than 25 words to be the most effective

In order to measure the effectiveness of the translatability guidelines, a focus group was asked to track the corrections they implemented on a text that had been processed by the customised CL checking tool *Sunproof* They found that the editing of CL texts was not much faster than non-CL texts but that the editors focused less on minor repetitive errors An interesting finding was that some sentences did not improve after CL rules had been implemented and, indeed, clarity was sometimes reduced in sentences where CL rules had been applied (ibid 2002)

> *some sentences that had been rewritten as advised by Sunproof did not improve clarity Further, in some rare cases, some sentences became even more ambiguous after revisions [sic]*

The reason for this is attributed to the fact that the writers had difficulty disambiguating their own complex sentences This is interesting as it suggests that the technical authors did not give much thought to the complexity of their sentences when they were writing the original document To conclude, Wells-Akis and Sisson make some significant claims regarding CL – they say that texts processed by Sunproof will be easier to translate and that the time required for translation will be reduced They also claim that human translation time is reduced and translation accuracy is improved It is unfortunate that no empirical evidence is offered to support these claims Nonetheless, Wells-Akis and Sisson offer an endorsement from the commercial world of the concept of translatability in a CL and MT environment

2.6.7 Conclusions on Translatability

In Chapter 1, we commented on the lack of intersection of CL rules and we suggested that some compromise might be necessary in identifying a suitable list of linguistic features to test in this study Of the seven rules identified in Chapter 1 as being "common" to the eight CLs that were analysed, five of the features they seek to control are included in Bernth and Gdaniec's (2001) MTranslatability list, i e rules governing the use of noun clusters, ellipsis, gerunds, active voice and relative pronouns The rule on sentence length that is shared by all eight CLs is also included in this list In addition, the explicit claim by Bernth and Gdaniec that their approach seeks to be TL- and system-independent makes their list of translatability indicators all the more appealing as a source for further study Bernth and McCord, Gdaniec, and Underwood and Jongejan all discuss the application of penalties in the context of specific software applications, to which the author did not have access during this study When combined, all of these factors led to the conclusion that Bernth and Gdaniec's list of translatability indicators would be the most suitable for the study at hand Although we acknowledge that our own study, for reasons that will be discussed in detail in Chapter 5, is limited to one SL-TL pair and one MT system, we feel that this more generic approach will be beneficial for researchers who may want to build on this study for other language pairs and/or MT systems

The term "translatability indicator" (Underwood and Jongejan 2001 363) is somewhat misleading since it gives the impression that it is a positive, rather than a negative, feature We therefore propose to improve clarity by changing the term to *negative translatability indicator*, or "NTI" for short

Bernth and Gdaniec's list of translatability "rules" is given in Appendix A There are 26 rules in all, corresponding to twenty-one easily identifiable linguistic (e g use of the gerund) or textual (e g use of footnotes) features Some minor adaptations were made to the list of NTIs For example, while there are three rules governing the use of the gerund, we counted "gerund" as one NTI For rule fourteen - *Avoid metaphors, idioms, slang, and dialect* – we chose to concentrate on just one of these categories (slang) because it was felt that the outcome for all four categories would be similar for MT output Finally, the last three rules, *Proofread and correct scanned documents*, *Avoid textual content in bitmaps*, and *Use mark-up wisely*, did not apply to the text in our study

2.7 CONCLUSION

The aim of Chapter 2 was to elaborate on the topic of Controlled Language, in particular in relation to Machine Translation and translatability assessment The primary aim of CL for MT - the reduction of ambiguity on different linguistic levels - was discussed We then addressed the notion of how the translatability of a text can be assessed Four approaches were compared and the conclusion was reached that Bernth and Gdaniec's (2001) generic approach to translatability measurement was the most suitable for this study The minor modifications made to their list of NTIs is also explained

Chapter 3

3. POST-EDITING

3.1 INTRODUCTION

There is evidence to suggest that the demand for more automation of the translation process is increasing The productivity increases reported following the introduction of translation memory tools in the 1990s (O'Brien 1998) appear to have plateaued, resulting in the recent creation of organisations such as the Translation Automation Users' Society (TAUS, Van der Meer 2003), whose ambition it is "to automate more than what until now seemed humanly possible"[25] Companies such as Symantec and Microsoft are now undertaking research on the possibilities and limitations of CL and MT (Roturier 2004) and localisation providers, such as SDL and VistaTech, report that their customers are now demanding post-editing services[26] Research on CL and MT is, therefore, gaining in importance However, the trend thus far has been to concentrate on the effectiveness of CL rules and the quality of MT output The all-important process of "correcting" MT output has been somewhat ignored This is despite the fact that Ryan's contention (1988 131) that

> *in an age when some systems can already translate over a million words in an hour, the time that it takes to run a translation becomes insignificant, the cost-effectiveness of the MT system must be measured largely by the effectiveness of the post-editing process*

still holds true some eighteen years on The idea that any post-editing effort expended must be taken into consideration when evaluating the cost-effectiveness of MT is central to this thesis Given the effort involved in implementing CL rules for MT, the onus on researchers to track the effort involved in post-editing MT output in CL scenarios seems particularly great To do so, however, we need to have a clear understanding of what post-editing effort is, and of how we can measure it These questions are dealt with in detail in this Chapter

In Section 3 2 we discuss the definition of post-editing Section 3 3 outlines research on the topic and gives a comprehensive overview of the most extensive study of post-editing to date (Krings 2001) Section 3 4 provides justification for research into post-editing by outlining the increasing demand for MT and by providing evidence of the success of MT/post-editing In section 3 5 we investigate the nature of post-editing and describe different types of post-editing as well as rules and guidelines that have been offered by various researchers The topic of computer-aided post-editing is also broached in 3 5 To conclude, 3 6 poses some questions about the skill-set required for post-editors and about training and education needs

[25] TAUS home page http //www translationautomation com (last accessed March 29 2006)
[26] Personal communication from both companies between 2004 and 2006

3.2 THE DEFINITION OF POST-EDITING

Allen (2003) says that the term "post-editing" is most commonly associated with a definition given in Veale and Way (1997) in which post-editing is understood as the "term used for the correction of machine translation output by human linguists/editors" The task of a post-editor, Allen goes on to say, is to "edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)" (Allen 2003 297)

According to several authors, the activity of post-editing differs from traditional translation and from revision, i e the activity of reviewing text translated by a human and correcting errors in content, spelling, punctuation, formatting, etc For example, McElhaney and Vasconcellos (1988 141) outline a number of parameters where post-editing differs from traditional revision

- The types of errors to be corrected are different

A reviser has responsibility for finding missing words, skipped passages, inadvertent repetitions, misspellings, mistakes in numerals, incorrect punctuation, inappropriate glosses and misconstructions of meaning In comparison, a post-editor has assurance that, at least on the mechanical level, no passages have been skipped, and that spelling errors are unlikely Pigott (1988 161) also holds that the errors corrected by post-editors are different from those corrected by revisers of human translation

- Misconstruction of meaning will occur at different levels

While misconstructions of meaning will occur in both human translation (HT) and machine translation (MT), they will occur at different levels Whereas misconstruction of meaning will frequently occur at the phrasal or lexical level with MT, misconstruction of meaning by a human translator will frequently result in the reconstruction of an entire sentence

- Post-editing seeks the minimum steps required for an acceptable text

The two processes of revision and post-editing are most alike, according to McElhaney and Vasconcellos, when the goal is to produce a text to be published This is when cohesion and style become significant factors However, even here they identify differences the post-editor will seek the minimum number of steps required to make a text acceptable and will therefore re-order less than a traditional reviser On the subject of differences between translating and post-editing, Senez (1998a no page numbers) says "A translator will always strive to disguise the fact that the text has been translated In the case of post-editing, it is enough for the text to conform to the basic rules of the target language, even if it closely follows the source text ”

Loffler-Laurian (1986, I 81) differentiates revision from post-editing by saying that the translator delivers what s/he considers to be a “final” version to the reviser, whereas MT output is never considered “final” Post-editing is, therefore, a process of *modification* rather than *revision* Loffler-Laurian (1985 71) also argues that machine translation and human translation should not be compared because they are such different activities In her opinion, the only valid comparison involving MT is one where two different MT systems are compared

Raw output from MT systems has become an acceptable medium for “information gisting” purposes For example, the European Commission, which has been using MT for many years, has found a new role for raw MT output in information gisting (Senez 1998a, 1998b) However, MT is also used to produce high-quality, publishable output and in this case post-editing is required It is expected that pre-editing the text in Controlled Language will reduce the post-editing effort However, it will not eliminate the need for post-editing, especially when the aim is for the target text to read as if it were written by a native speaker of the target language Consequently, when implementing a combined CL/MT solution, one has to also consider the topic of post-editing

Unfortunately, the literature available on post-editing rules and techniques is somewhat sparse, as Allen (2002 28), in his review of Krings (2001) reminds us

> *Krings states that the problem in the past is that there has been relatively little effort spent on studying PE This 1994 statement still appears to be valid, unless there is a lot of hidden PE project work that is not being made publicly available*

Allen and Hogan (2000 64) claim that very few reports are publicly available on post-editing and they go on to suggest that most translation houses involved in post-editing MT are re-inventing the wheel They (ibid 62) explain how post-editing and CL fit into the translation process

> *Authoring/Translation Workflow now links some of the traditional tasks of technical writing and translation with issues of ontological organization, electronic communication, electronic delivery, informational (sic) retrieval, and data-mining, to the extent that the notion of CL is beginning to encompass a range of tasks in the areas of authoring, editing, translation, translation editing/revising*

For them, post-editing is seen as “another type of CL processing” (ibid 63) where the post-editor works on a “tri-text” composed of source language, MT output and post-edited material and where the modifications made by translators/post-editors are similar to the modifications writers are encouraged to make on their source texts

3.3 Research on Post-Editing

The most extensive study on post-editing carried out to date is by Hans P Krings (Krings 2001) This research was carried out during the early 1990s at the University of Hildesheim and the English translation by Koby was published in 2001 Krings's hypothesis-finding study compares the mental processes involved in post-editing machine translation with those of translation In taking stock of the literature on post-editing, Krings observes a striking research deficit regarding virtually all questions related to post-editing Most literature published on the topic at the time included unsystematic observations and subjective evaluations (e g Loffler-Laurian 1984, 1986 parts I and II, Wagner 1985, 1987) [27]

The only work known to Krings where post-editing effort was recorded in terms of the time required was a *Diplomarbeit* (thesis) (Sander 1990, University of Hildesheim) where a comparison of fast post-editing, complete post-editing and conventional translation was undertaken The conclusion from this study was that the task of full post-editing does not lead to a decrease in the time expended in producing the target text, but rather an increase of between 10 and 60% Partial post-editing can save between 10 and 65% of time, if only the human effort and not the computer processing time is included

Additional interesting findings from Krings's own empirical research include the fact that reference works are consulted more frequently during translation than during post-editing, leading to the hypothesis that MT can reduce reference work usage Also, the task of post-editing with a source text requires more source text-related processes than translation [28]Krings concludes from this that the text comprehension processes necessary in post-editing cause additional processing effort A very interesting finding is that, when measured by the number of source text-related processes, post-editing effort is highest not for poor quality MT, but for medium quality MT

Krings examines the number of "focus changes" that occur in each of the tasks under investigation [29] It is assumed that the more often a post-editor has to change focus, the greater the cognitive effort will be Since the post-editor has not two (ST and TT), but three texts (ST, MT output and final TT) to focus on, Krings asks if this increases the cognitive effort and consequently reduces any perceived benefits of using MT and post-editing When counting the number of focus changes, he weights them according to verbalisation volume because a subject who verbalises more while thinking aloud changes focus more too The

[27] Several undergraduate theses on post editing exist e g Sander (1990), Falkenheimer (1992) Gennrich (1992) Mohr (1992) and Zucko (1993) Unfortunately the libraries where these studies are stored do not allow theses to be removed with the result that this researcher was unable to consult them directly Nevertheless, Krings (2001) includes a detailed commentary on all of the above

[28] Krings's 2001 study was a process oriented one This is discussed in more detail in Chapter 4

[29] Focus changes were identified via think aloud protocols produced by the research subjects The topic of TAP will be discussed in Chapter 4

results show that the number of changes in attention focus is almost twice as high in the 'post-editing with source text task' as in translation. The value for post-editing without source text is intermediate, leading Krings to conclude overall: "the results of this study show no appreciable reduction in the cognitive effort needed to produce a target text by using a machine translation" (ibid: 320).

Of most significance to this study, Krings poses the question: which factors determine post-editing effort? Krings states that "the question of post-editing effort is the key issue in the evaluation of the practicality of machine translation systems" (ibid: 178). The term "post-editing effort" requires further differentiation, however. Three different types of post-editing effort are identified: temporal, cognitive and technical. The only study known to Krings where the technical effort required for post-editing on-screen was considered was Gennrich (1992). In this case, "technical effort" is defined as the "physical operations that must be performed on a machine translation to correct any errors or defects" and this includes deletions, insertions and rearrangements. Technical effort, however, is just one externally observable aspect of post-editing. The critical, decisive variable, according to Krings, is the quality of the MT. Wilms (1981: 40) supports this by claiming that post-editing effort is inversely proportional to MT quality. Krings also mentions Sander's efforts to find a relationship between text-type and post-editing effort. However, results provided no indication of a relationship and Sander concluded that it is the syntactic complexity of the individual texts that is the deciding variable in determining the quality of MT and, consequently, the post-editing effort. Krings poses a question regarding which relationships exist between the translating or post-editing process and the translated or post-edited product. According to him, the most important discovery of translation process research to date is the existence of a process/product ambiguity, i.e. "linear conclusions about process based on product are inadmissible" (ibid: 176).

Temporal post-editing effort is the most visible and is considered the most important economically. It is also the most easily measured. The broader issue of determining variables needs consideration when measuring temporal post-editing effort. These variables include the extent and type of deficiencies found in machine translation. Krings points out that correction effort is not equal for all deficiencies.

The determining variable underlying temporal post-editing effort is cognitive post-editing effort (ibid: 179). This refers to the extent and type of cognitive processes that must be activated to remedy a given deficiency. Cognitive post-editing effort cannot be observed directly. Therefore, according to Krings, TAP (think-aloud protocol) should be used to investigate it.

Technical post-editing effort refers to the deletions, insertions and rearrangements the translator performs when post-editing the MT output. The number of cursor movements needed as well as the deletion and insertion operations may serve as a direct indicator of the technical post-editing effort. While a deficiency in MT may be easily recognised and a correction strategy easily developed, its technical implementation may not be correspondingly simple: “It is precisely in on-screen post-editing that cognitive post-editing effort and technical post-editing effort seldom converge” (ibid: 179).

Krings suggests that post-editing effort must also be differentiated using another criterion, that of comparison. One possible measure of comparison is with *fully-automatic high quality machine translation*, i.e. MT output that needs absolutely no post-editing. Krings terms this measure *absolute post-editing effort.* It is an indicator of the difference between machine translation and human translation and it “must be included as a corrective factor in research and development efforts” (ibid: 180).

Krings proposes another standard measure to evaluate the practicality of machine translation, which is called *relative post-editing effort.* This is a measure of the effort required for machine translation plus post-editing, compared with human translation. The connection between absolute post-editing effort, relative post-editing effort and translation effort can be represented in the following formula:

$$PE_{rel} = \frac{PE_{abs}}{TE}$$

“PE_{rel}” is the relative post-editing effort, while “PE_{abs}” is the absolute post-editing effort and “TE” is the translation effort. Relative post-editing effort may have any value between zero and infinity. Zero means perfect MT and no post-editing. However, a certain temporal value is always required to even ascertain if MT output requires post-editing, so an absolute value of post-editing effort of “0” is very much a hypothetical value. A value between zero and one means that post-editing effort was less than translation effort. A value of one means that translation and post-editing effort are equal and a value greater than one means that post-editing is more time-consuming than translation. For example, if the PE_{abs} value was 20 minutes for a given text and the TE value was 30 minutes, PE_{rel} would then be 0.66, which means that post-editing effort was lower than translation effort.

The decisive difference between absolute and relative post-editing effort lies in the fact that the latter reveals how efficient an MT system is. Krings cautions against making the assumption that low post-editing effort indicates a good machine translation. In fact, if the post-editing effort is low, it is likely that the human translation effort would have been low too, thanks to the quality of the source text Detailed investigations of very high or very low post-editing efforts must be used to draw connections with source text characteristics According to Krings (ibid 182)

> *The goal of the investigation must be to extrapolate the results achieved to other forms of source language material, with a long-range objective of forming a reliable prognosis of post-editing effort This prognosis may then serve as the basis for a well-founded decision between the alternatives of human translation versus machine-assisted translation*

Krings also examines post-editing effort from the point of view of textual similarity, i e "surface lexico-syntactic similarity between the target text and the machine translation" (ibid 292) A surprising finding was that the similarity factor was almost the same for the task of 'post-editing with the source text' and that of 'post-editing without the source text' Despite this, in his process analysis, Krings found that there were significant differences in these two processes In his opinion, this points to a frequently mentioned ambiguity between process and product The explanation put forward by Krings for this phenomenon is that

> *the objective defects in the machine translation lead to different strategies on the process level, which, however, lead to essentially the same corrections on the product level, independent of the availability of the source text*
>
> *(ibid 298)*

Krings observes that even for MT rated as "good", a substantial number of changes are made during post-editing He offers two possible explanations for these figures (1) there are different severity levels of MT defects, and less significant errors such as syntax errors result in high evaluation ratings but still need considerable changes on the sentence structure to make a well-formed sentence and (2) post-editors may make more changes in the machine translation process than absolutely necessary

One of Krings's main findings is that the process of post-editing is "very similar" to that of translation (ibid 557)

> *the analyses showed that normal post-editing (that is, post-editing with source text) is a task very similar to normal human translation, both qualitatively with regard to the types of cognitive processes occurring, and quantitatively with regard to the frequency distribution of such processes*

Like Krings, Olohan (1991) examines post-editing from a cognitive viewpoint, using think-aloud protocols Her source text was a user manual for a toaster and was written originally in German Olohan used the METAL MT system to translate this into English Subjects post-edited a "mixed" file, i e each target language segment was followed by a source language segment Four native speakers of English, two students and two professional translators, were asked to post-edit the raw MT output A syntactic analysis was carried out on the individual post-edited translations of the 300-word source text, and Olohan identified three major syntactic changes that occurred during post-editing active to passive, noun to verb and inter-translation unit changes

Her observations on active to passive changes are

- In many cases, the passive voice is "used" but not "chosen" Its use is influenced by sentence structure, clause order and choice of subject/topic (Olohan ibid 67)
- In those cases where the passive voice was selected as an alternative to active voice, subjective preference or knowledge of syntactic norms of this particular text type seem to have formed the basis for this selection
- Certain syntactic changes are made purely on the basis of personal preference and a subjective judgement that these syntactic forms are stylistically superior

In her analysis of Noun to Verb changes, Olohan examines individual nouns, noun phrases and prepositional phrases (N + PP) that are changed to a verb in the English post-edited version She concludes that, when analysing TAPs, noun to verb changes are more enigmatic than active to passive voice changes

> *The progression from the often ungrammatical MT to a satisfactory subordinate clause, gerundial phrase or infinitival construction was frequently effected without perceptible intermediate changes in the TAPs* [30]
>
> *(ibid 143)*

Olohan distinguishes between two types of change those which occur as a result of native speaker competence and those which are prompted by contrastive knowledge of the two languages involved in her study (ibid 148)

"Inter TU" changes are defined by Olohan as "syntactic changes, the relevance of which transcends the boundaries of the TU(s) in which the change is made" (ibid 153) They include the introduction of linguistic elements with a referential function to link two TUs, or the re-ordering of elements to influence thematic structure In this regard, Olohan examines cohesion, coherence and intertextuality and notes that there are significant differences between participants in the number and type of changes made to influence cohesion and coherence and that subjects implemented cohesion changes above the TU level (ibid 183)

Olohan's main conclusions are

- All subjects demonstrated what she calls the "Never Mind It'll Do" syndrome (ibid 192) She terms this strategy a "reduction strategy" and comments that subjects

[30] This is further evidence of the phenomena of "automatisation" which is discussed in detail by Krings (2001) and is addressed in Chapter 4

may well have been influenced by the knowledge that they were participating in an experiment and that the text would not be published

- Post-editors can post-edit without actually understanding the source text or the raw MT output and, yet, still produce a version of the target text that is correct (ibid 200)
- Subjects can arrive at similar products in very different ways (an echo of Krings's assertion on product/process ambiguity) Some of Olohan's participants worked directly with the raw MT text while others looked at the source text first without even considering the usefulness of the MT output

Olohan's study brings to light some necessary skills for post-editors and these will be addressed in the section on education and training below

Loffler-Laurian (1986, parts I and II) carried out an analysis of two types of post-editing – rapid and conventional – with English as a source language and French as a target language and drew some conclusions about their linguistic features[31]

- Rapid post-editing uses comprehensible, but more banal or common vocabulary while conventional post-editing uses more precise or specialised vocabulary The decision to use banal or more specialised terminology is dictated by text function, according to Loffler-Laurian
- Conventional post-editing uses more fixed phrases than rapid post-editing Loffler-Laurian suggests that it is the literal nature of rapid post-editing that suppresses the use of fixed phrases In conventional post-editing, many of the lexical changes involve conjunctions, determiners or modifiers Loffler-Laurian terms these "faux-semblants stylistiques", which means that the post-editor utilises quasi-synonyms to create a higher register in the text (ibid 87)
- Conventional post-editing prefers verbal formulations to nominal ones and this is especially true in lists (ibid 88)
- An explicit agent is more common in conventional post-editing than in rapid post-editing
- In conventional post-editing, ideas are more likely to be expressed in the positive while in rapid post-editing ideas are more likely to be expressed in the negative

Unfortunately, Loffler-Laurian provides little detail about the amount of data used or the number of subjects in her analysis, or her methodology While her observations on post-

[31] We will discuss the different "types" of post editing in more detail later in this Chapter

editing are helpful, we should keep in mind that some of the linguistic features observed might be a function of the source and target languages or of the subjects themselves Loffler-Laurian herself observed different post-editing tendencies between anglophones and francophones, where the latter had a tendency to make logical relations in a sentence more explicit by adding conjunctions or punctuation (1985 72) Senez (1998a) also observed different post-editing techniques from one language pair to another, but she does not elaborate on this

Some evidence of research on post-editing can be found on the World-Wide Web too For example, an abstract entitled "User Behaviour During Post-editing in Machine Translation" appears on the Information Processing Society of Japan web site [32] Unfortunately, this paper is accessible to members only A project undertaken by the Natural Language Group of the Information Sciences Institute at the University of Southern California provides some explanation as to what is involved in post-editing [33] Finally, it is surprising to note that a book published relatively recently on the topic of revising and editing for translators (Mossop 2001) includes only two pages on the topic of "Revision of Machine Translation"

3.4 THE DEMAND FOR POST-EDITING

More than twenty years ago, Van Slype (1982 80) estimated the market for rough translation without post-editing to be at 30% of the actual market for formal, written, human translation He maintained that for translators to be willing to accept raw MT for post-editing, the correction ratio would have to be lower than 20%, i e one correction in every five words (ibid 80) If Controlled Language were added to the formula, the expected correction ratio would be 5-10% At that time, Van Slype estimated that MT with post-editing was cheaper than human translation His estimated cost for human translation in a public institution was £8 54 per 100 words of source text and £2 67 in a private translation bureau MT, on the other hand, was estimated to cost £6 07 in a public institution and £2 40 in a private bureau This included costs for data capture, translation, typing and post-editing, but excluded software investment costs (ibid 88)

[32] ARAI, Yoshinori, SHIMADA Atsuo NARITA Masumi IPSJ SIGnotes Natural Language, No 0 84, http //www ipsj or jp/members/SIGNotes/Eng/01/1991/084/index html (last accessed on 03 November, 2002)
[33] http //www isi edu/natural language/mteval (last accessed 09 November, 2002)

Machine Translation was implemented at the European Commission as early as 1979 Demand, however, was low in the beginning Senez (1998b) attributes this to a number of factors limited availability of texts in electronic form, lack of information, relative instability of the e-mail system, lack of awareness about MT Improvements in the e-mail system and a vigorous promotion campaign resulted in a dramatic growth in the number of pages being machine translated, which grew from 30,000 in 1990 to 200,000 in 1997 (ibid 289) In 2001 the European Commission machine translated over 750,000 pages [34] This figure rose to over 860,000 pages by 2005 [35] The increase in demand has resulted in a rapid post-editing service being set up by the translation service of the European Commission Fifty per cent of requests for MT at the Commission concern combinations of French and English, while 40% concern combinations of German with French or English, and 10% of the requests are for Italian, Spanish or Portuguese in combination with French or English (ibid 291)

Allen (2003) attributes a recent growth in demand for post-editing to increased globalisation activity According to him, the majority of post-editors work for PAHO, the European Commission, or Caterpillar, as well as some translation or localisation companies such as the Detroit Translation Bureau Other localisation companies who have recently introduced machine translation in-house could be added to this list, e g Lionbridge and SDL International [36]

More recent public comments demonstrate that there is indeed a growing demand for MT For example, it was predicted that the MT market would grow from 30 million euro in 1998 to 200 million euro a few years later (Goukens 1998) In the year 2000, it was reported that the AltaVista Babelfish site was used one million times a day (Bennett 2000) A large proportion of the demand for MT via translation portals could of course be for information gisting purposes only It is, however, safe to assume that some users of information translated automatically via translation portals will decide to have some proportion of the translated information post-edited It is, unfortunately, difficult to capture figures depicting this demand, but it is reasonable to assume that with a growth in demand for MT, there will also be a growth in demand for post-editing

3.4.1 Evidence of Success

> *“ in the end an MT system will stand or fall depending on the human environment in which it is placed, and [ ] some of the most important factors cannot be measured”*

[34] Cameron Ross European Commission personal communication November 2002

[35] ECMT Helpdesk personal communication January 2006

[36] This information was acquired by the researcher at the Society for Automotive Engineers Multilingual TOPTEC Symposium in Nashville, Tennessee, 3 4 October, 2002 In addition, the growing demand for post editors in the localisation industry was confirmed by Bert Esselink (Lionbridge) in an oral presentation given to translation students at Dublin City University in December 2005

> *(Vasconcellos and León, 1985 135)*

Evidently, PAHO's implementation of MT is successful since Vasconcellos and Leon report that a fully trained post-editor, working on-screen, produces standard quality output two to three times faster than traditional translation allows (4,000 – 10,000 words per day) (ibid 122)

Several others have published evidence of success with MT and post-editing For example, Pigott observed a doubling of translator output within a few months of post-editing MT (1982 65) and later records a four-fold increase in the rate of translation and a halving of costs

> *Although no official statistics are available, it would appear that efficient use of the system can speed up the rate of translation by as much as a factor of four, and that it can reduce overall costs by at least half*
>
> *(Pigott 1988 163)*

He adds, however, that many translators report little or no increase in throughput and this is attributed to lack of experience with post-editing, poor quality of the source text, deficiencies in the dictionary and psychological reactions to MT and to word processing [37] Pigott notes that it is only with experience that translators begin to appreciate the benefits offered by MT

Reporting on the implementation of Systran at General Motors in Canada, Sereda (1982 120) notes that post-editors can work at a rate three to four times faster than "manual" translators Factors influencing the effectiveness of MT and post-editing are listed as the linguistic performance of the system, the source language to be translated, and the availability of terminology

Wagner (1985 205) also records a rate for post-editing that is four times faster than manual translation However, in the same article, she also reports that many EU translators saw Systran as a hindrance rather than a help, because it limited their freedom of expression (ibid 213) In her description of the implementation of EC-Systran, Senez (1998a) claims that the average price of a post-edited page is slightly more than half the average price of a page translated by freelance contractors working for the Commission's translation department

A recent report on best practices in post-editing (Joscelyne 2006) reported average daily throughput rates for publishable quality post-editing as being approximately 5,250 words per day [38] The average translation throughput expected in the translation industry for

[37] Pigott was writing at a time when word processing was new for most translators

[38] Joscelyne draws on reports from companies who have implemented MT The rates were reported in terms of pages per day, which is somewhat ambiguous (how many words or characters are on a page?) For comparison purposes, I have converted the rates to words per day, assuming an average of 350 words per page

the text type in this study is 2,000 words per day.[39] Again, this suggests that MT plus post-editing can deliver faster throughput rates than human translation unaided by translation technology.

We can see that there is evidence that MT and post-editing can be faster and cheaper than manual translation. However, this evidence has not provided an impetus for an increase in the use of machine translation for numerous reasons: First of all, the evidence is anecdotal - few details are given regarding how figures of "three to four times faster" or "half the average price" are arrived at. Then, success depends on factors which are difficult to measure: as Vasconcellos and León (1985) and others have stated, the success of an MT system depends largely on human factors that are very difficult to measure, for example, the attitude of translators or managers to MT technology in general. The successful implementation of MT in the future depends not only on technological advances in Natural Language Processing (NLP) techniques, but also on the training translators receive. We return to this topic below.

3.5 The Post-Editing Process

3.5.1 Types of Post-Editing

McElhaney and Vasconcellos (1986) differentiate between *batch* post-editing and *interactive* post-editing. Batch post-editing involves manipulating the text after a full machine translated version has been produced. Interactive post-editing, on the other hand, involves manipulating the text as it is being translated. This process usually involves prompts from the machine translation system for clarification when some ambiguity has been encountered. Somers (1997) suggests a variant of the interactive approach. He talks about "post-editing" the source text. This involves evaluating the MT output before any post-editing takes place and editing the source text to improve the raw MT output. It is important to differentiate between the application of controlled language rules to a source text and Somers's suggestion for post-editing the source text. The latter does not involve the application of any rules, but simply the transformation of the source text, possibly even into ungrammatical utterances, with the aim of producing a better target text. Somers points out that there are, as yet, no metrics for comparing the efficiencies of post-editing the input to MT as opposed to the output from MT.

Bédard (1992) also proposes an alternative to post-editing which he calls "prétraduction", or pre-translation. Pre-translation is defined as:

[39] Personal communication, Phil Ritchie, CTO, VistaTech (April 2006).

> *une* traduction partielle *du texte, par simple remplacement global d'une partie des mots du texte au moyen d'un dictionnaire d'equivalences*[40]
>
> *(ibid 745)*

According to Bedard, this saves time *in* terminology look-up and typing and eliminates the time required to understand sometimes incomprehensible MT output He offers anecdotal evidence to suggest that this method is (a) more economical and (b) more acceptable to translators (ibid 743/745) However, he does not provide convincing evidence for how he arrived at this conclusion

Loffler-Laurian (1986, I and II) differentiates between *rapid* and *conventional* post-editing Although these two types of post-editing can be differentiated along a time line, Loffler-Laurian argues that is better to differentiate them along linguistic lines Another criterion for differentiating the two is the level of "bruit linguistique" or "linguistic noise" that can be tolerated in a text (ibid 226) Linguistic noise can be caused by the extraordinary use of a word, unexpected word order, unexpected concordance etc Loffler-Laurian reports that studies have been carried out to determine the impact of linguistic noise on comprehension in spoken language domains, but at the time of writing her article no such studies existed for written, specialised language

Senez (1998a) differentiates between "correcting MT" and "post-editing" where correction means production of a final, flawless product According to her, rapid post-editing is a perfectly viable option as long as three criteria are met (1) the customer needs the text urgently, (2) the text is not destined for publication, (3) the customer makes an informed decision and weighs up the advantages of a faster service against the risk of loss of quality (1998b 292)

Rossi (1982) identifies three types of post-editing The first type limits post-editing to making the text comprehensible The second type strives to make the text as authentic as possible in the target language and the third type adapts not only grammar but also style and content to the target language

Vasconcellos and Leon (1985) and Vasconcellos (1986a) maintain that the degree of post-editing is determined by the purpose of the translation, the user's editing resources, the time frame and the structural linguistic considerations in the text itself Wagner (1985), on the other hand, maintains that the degree of post-editing is determined by the individual's preferences and the quality of raw MT output Allen (2002) attributes the degree of post-editing to several, similar factors, i e the user requirements, the volume of documentation to

[40] "a partial translation of the text using simple global replacement of some of the words by means of a dictionary of equivalences" (my translation)

be processed, quality expectations, turn-around time, perishability of information and text function He differentiates between "inbound" and "outbound" MT where the former involves translation for comprehension and the latter translation for communication For inbound MT, one may decide not to do any post-editing or to do a rapid post-edit For outbound MT, again one may decide not to do any post-editing or to do a rapid or "partial" post-edit or a "full" (conventional) post-edit

Post-editing can occur on-screen or on paper Since most translators now work with word-processors, on-screen post-editing is more usual Vasconcellos claims that post-editors can work faster on-screen, especially if they are equipped with good key-boarding skills (1985 120) Vasconcellos also identifies a "functional" approach to post-editing where the post-editor works from left to right, backtracking as little as possible This not only saves time, but is more consistent with the natural production of text (1986a 142)

3.5.2 Post-Editing Rules

In 1985, Wagner suggested some guidelines for post-editors, which are reproduced in Allen 2002

- retain as much raw translation as possible
- don't hesitate too long over a problem
- don't worry if style is repetitive
- don't embark on time-consuming research
- make changes only where absolutely necessary, i e correct words or phrases that are (a) nonsensical, (b) wrong, and if there's enough time left, (c) ambiguous

Loffler-Laurian also suggests guidelines for post-editing (1986, II 227) Her terms of "obligatory" and "necessary" post-editing refer to the fixing of proper nouns and abbreviations, technical terms or lexical units, expressions of relations in a sentence, and common phrases (and their usage) "Non-justified" post-editing, on the other hand, involves replacing accurate terms with synonyms, changing tense, aspect, or voice, changing word order, and making implicit relations explicit In an analysis at the European Commission, Loffler-Laurian found that obligatory changes were carried out by all post-editors, necessary changes were carried out by 50-75% of post-editors, and unnecessary changes were carried out by less than 25% of post-editors (1985 73)

Loffler-Laurian also proposes principles for post-editing which are divided into three categories psychological and linguistic attitudes, general rules and specific rules Under psychological and linguistic attitudes, Loffler-Laurian reminds the translator, inter alia, that

MT does not take away one's freedom but that, rather, it should be viewed as an aid Her general rules specify that the post-editor ought to keep as much of the raw translation as possible and that s/he ought to replace what she terms the "3Is" with the "3Cs", i e replace that which is "Infidele" (unfaithful) and/or Incorrect and/or Incomprehensible with that which Conforms, is Correct and Comprehensible Her more specific guidelines draw attention to terminological and punctuation errors, as well as logical relations and tense, among other things

3.5.3 Error Classification

Green (1982 101) defines a post-editing error as "any feature of the translation which causes the post-editor to put pen to paper" He proposes three categories of error minor, major and grey areas A minor error includes the misuse or omission of the definite article, incorrect preposition or personal pronoun or the wrong choice in translation – usually of a noun - when alternatives are possible According to Green (ibid), minor errors are easy to identify and post-edit

Examples of major errors include a literal translation of idiomatic expressions or a part-of-speech error Green suggests that the quickest remedy for these types of errors is to delete the clause or sentence and write in one's own translation from scratch

"Grey areas" are described as "doubtful translations or "near misses" (ibid 102) In such cases, the post-editor must decide whether or not to alter the text This decision will be influenced by many of the factors discussed previously, e g time, life-cycle of the text and so on

Senez (1998a) also talks about errors in terms of their need for alteration With experience, post-editors gain a feeling for errors and whether or not they require modification, or in Senez's words, they gain a feeling for "that which must be changed, that which may be changed and that which is superfluous to change" (ibid no page numbers)

In his survey of attitudes to post-editing in the European Commission, Lavorel (1982) identified four categories of errors In order of seriousness they are incorrect verb forms, mistranslation of prepositions, literal rendition of common idioms, consistent translation of a word in one manner when the context demands another translation

Reporting on the use of machine translation at General Motors in Canada, Sereda (1982) notes that different types of errors have different effects on the post-editing process

> *Minor errors involving articles and verb/adverb rearrangement can be resolved quickly and easily by the translator On the other hand, certain kinds of structural errors can be extremely difficult to correct, perhaps requiring the complete rewrite of the affected sentence*
>
> *(ibid 120)*

Pigott (1982) also alludes to minor and major errors, saying that an average sentence could require up to four or five changes, many of them minor, but in some cases whole sentences or parts of sentences are retranslated

Loffler-Laurian (1983) provides the most detailed account of post-editing errors An error is "tout ce qui a donne lieu a une modification de la part du post-editeur" (ibid 66) [41] She noticed that sometimes text is modified when no error occurs and sometimes real errors are ignored The latter cases are, however, rare Her analysis was carried out in the MT environment of the European Commission in Luxembourg with the aim of identifying text types that were suitable for machine translation and post-editing The objective was

> *á savoir de distinguer avant la mise sur machine ceux des textes qui sont susceptibles d'être "bien" traduit et ceux qui auront besoin d'un nettoyage tel que la traduction humaine aurait ete plus rapide et plus economique que la traduction machine suivie de l'etape correctrice*[42]
>
> *(ibid 65)*

In other words, she wanted to be able to identify texts with a high machine translatability index The methodology involved an examination of types of post-editing errors rather than an examination of source text characteristics

Loffler-Laurian bases her analysis on text-type linguistic categories [43] Twelve sub-categories were chosen based on how frequently certain features occurred in her corpus The categories are

1 Vocabulary/terminology

2 Abbreviations and Proper Nouns

3 Prepositions

4 Determiners (articles and demonstratives), verb modifiers

5 Verb forms (tense)

6 Voice (active/passive) and personalisation (e g by using il or on)

7 (Non-) Expression of modalities

[41] "everything which requires a modification by the post editor" (my translation)

[42] "to know prior to machine translation, how to distinguish those texts that could be translated "well" from those that would require such an amount of post editing that human translation would have been faster and more economical than machine translation followed by post editing (my translation)

[43] She defines these as semantic and syntactic categories which exist at least in all European languages (ibid 67)

8 Negation

9 Logic relations

10 Information added or deleted

11 Word Order

12 General problems with morphology or multiple determiners, which can usually be linked back to poor syntactic analysis

Errors could be classified into three major categories single word errors, errors of relation, and structural or informational errors The analysis also had three categories of occurrence no occurrence, at least one occurrence, and at least five occurrences For single word errors, the word category was recorded (e g proper noun, adverb, article etc) as well as the type of problem (e g nonsense, lack of precision, alteration of style etc) The category "errors of relation" included errors in uses of tense, modality, personal/impersonal, negation/affirmation etc Structural or informational errors included problems such as repetition, incorrect logical relations and word order

Loffler-Laurian's analysis involved recording errors and the types of changes made She then created a typology of errors and corrections and classified documents according to this typology To make good use of this, Loffler-Laurian states that it would be necessary to train people to perform a rapid classification of text types according to their suitability for MT

In his research on post-editing, Krings (2001) also devised an error classification system according to the following criteria

- Lexical Part-of-speech recognition error
- Lexical Other
- Morphology Word formation
- Morphology Other
- Syntax Word Order
- Syntax Other
- Stylistic usage norms
- Punctuation
- Textual Coherence
- Textual Pragmatics
- Literal transfer from ST

If there were multiple errors in any sentence, all errors were counted as discrete errors Krings collated a number of sample sentences for further analysis These were categorised into "poor", "medium" and "good" quality He found a correlation between the evaluations of the translators participating in his study and the number and type of MT errors An analysis of the ratio of number of errors to number of words in each of the quality categories (poor, medium and good) reveals that there is not much difference between the three - the ratio for sentences classified as poor quality was 1 2, for medium quality it was 1 3 and for good quality it was 1 4 This leads Krings to suggest that in a text of 360 words, one could expect 90 MT errors if the output was, on average, "good"

Although there is not a large difference in error count for each of the three quality classes, Krings found later in his analysis that the difference in error count exerts a considerable effect on the post-editing process (ibid 267) An additional discovery was that there is not a linear but rather an exponential relationship between the number of MT errors and the difficulty of post-editing (ibid 267)

Regarding the type of errors that occurred, Krings relates that lexical errors occurred in all sentences, but syntax errors did not occur in all sentences All part-of-speech (POS) errors occurred in those sentences rated as "poor" The latter observation leads Krings to the hypothesis that POS errors are particularly critical to MT quality and this is further endorsed by his analysis of 'post-editing without source text' where this type of error frequently results in a failed repair of the text (ibid 269)

Of the sentences in Krings's study that were post-edited without reference to the source text, the quality improved in forty-six out of fifty-two cases This is a success rate of 79% [44] The success rate for repairs was highest for word order and punctuation errors (94%) and this was followed closely by the categories of stylistic usage norms (92%), text coherence (90%) and morphology (88%) Syntax, lexical and text pragmatic errors had a 75% success rate and POS errors had a comparatively low rate of 54% There were a small number of errors that could not be corrected at all

3.5.4 Computer-Aided Post-Editing (CAPE)

Some organisations were quick to recognise that the clever use of a word processor could speed up the post-editing process (Wagner 1985, Vasconcellos 1986a) More recently, there have been suggestions that an "automatic post-editing tool" is viable (Knight and Chander 1994, Allen and Hogan 2000)

[44] Krings does not provide comparable figures for the task of post editing with the ST

In her 1986a paper, Vasconcellos describes how macros designed specially at PAHO aid post-editors in their daily tasks Two types of macros are used one that moves chunks of text and one that addresses particular constructions, for example, deletion of the next occurrence of "the" in the output translation (ibid 13)

> *Source Text* La teileriosis es transmitida por las garrapatas
> *MT output* Theileriasis is transmitted by the ticks
> *After Macro* Theileriasis is transmitted by ticks

Another example is the automatic transformation of a noun phrase from "(the) N1 of (the) N2" into "N2N1" or "N2's N1" (ibid 138) According to Vasconcellos, the use of macros involves a multiplicity of considerations

> *for example, whether or not the phrase is a common collocation in English, whether the discourse is formal or designed to convey a sense of "shop talk" or those "in the know", whether a head noun is an action, state, or process, and, perhaps most important, how the information is distributed with regard to the preceding and subsequent text*
>
> *(ibid 138)*

Santangelo (1988) also gives details on how computers can aid the post-editing process PAHO's "Engspan" system, for example, was capable of flagging a term to indicate that it had been researched and retrieved to the correct position in the sentence

Wagner (1985) reports on the use of automatic word processing functions, such as search and replace, to help speed up the post-editing process Somers (1997 202) talks about the "systematicity" of post-editing and how a word-processor's features can be useful However, he warns of the limitations of the "search and replace" function until "linguistically intelligent word-processors" are available

In 1994, Knight and Chander discussed the possibility of building "detachable post-editors" which could be used across different MT systems They envisioned two types of post-editing *adaptive* and *general* The idea behind an adaptive post-editing module is that it would monitor what errors were corrected by human post-editors and how they were corrected and it would then begin to emulate this process One could start this process off by using a corpus of "pre-postedited" text and post-edited text and then use statistical MT methods to learn the mappings A general post-editing module, in contrast, would be used to aid the correction of text generated by a variety of MT systems Presumably, one could have a general post-editing module that is also "adaptive"

Allen and Hogan (2000) propose an automatic post-editor (APE) which allows them to extract post-editing changes from "tri-text", i e corpora with source text, MT output, and post-edited text They carried out some research funded by the European Commission's "Service de Documentation" where the objective was "to identify the most frequent constructions [...] and to allow the APE module to learn the corrected forms from the post-edited versions of the tri-texts" (ibid: 67). By doing this, they argued, it would be possible to automatically learn what changes post-editors have applied to texts and then to develop a set of post-editing rules. They conclude that APE is a viable concept and that, with a significant amount of *additional training* and testing, it would be possible to reduce the number of post-editing fixes a human has to implement.

Povlsen and Bech (2002) also use the term "APE" to refer to their automated editing environment for output from the PaTrans MT system. In a survey by them, word order problems were ranked in first place as being most problematic for the English-Danish language pair and their automated editing environment tackles this problem. They claim to have recorded performance improvements of between 10 and 15% as well as improvements in cost efficiency. However, they do not offer any information on how cost efficiency was measured.

Sayáns Gomez and Villar Conde (2003) discuss the use of a post-editing toolbar for Spanish to Galician machine translation. However, the toolbar they describe has more to do with marking and labelling errors than automatically fixing them.

Joscelyne (2006: 8) reports that Symantec is investigating the use of pattern matching tools at the "post-processing stage" to "catch frequent mistakes and speed up the post-editing process". The features that Symantec are interested in at the post-processing stage include capitalisation, incorrect spellings, missing contractions, word order, formatting and inconsistent punctuation.

3.6 Training and Education for Post-Editors

3.6.1 The Need for Post-Editing Training

Krings (2001) considers how valuable a comprehensive theory of post-editing processes might be for the teaching and acquisition of post-editing competence. In his opinion, the results from translation process research will have direct significance for the planning of suitable teaching methods. Krings maintains that "it seems to be just a matter of time before post-editing of machine translation will also become a component of translator training to one extent or another" (ibid: 178).

Vasconcellos (1986a: 145) maintains that post-editing skills are developed gradually. The level of comfort with post-editing is greatly increased after 100,000 words (1 month of full-time post-editing). Somers (1997: 201) also reports that it is recognised by many that post-editing is a skill that needs to be "honed" Companies wishing to implement machine translation technology would therefore benefit if translation graduates were already "comfortable" with post-editing Additionally, post-editing skills would give translators an extra boost when it comes to finding employment opportunities

Translators who do not have post-editing skills are frequently hostile to machine translation technology Common arguments against MT include a dislike for correcting repetitive errors that a human translator would never make, a fear of losing language proficiency by working with poor MT output, and a dislike of having one's freedom of expression limited (Wagner 1985 213) However, translators who embrace post-editing report that their day-to-day work becomes much more interesting [45] Drawing on her experience of implementing Systran, Ryan (1988) maintains that the more and the earlier the translator is involved with the implementation of machine translation, the faster a usable system can be developed Senez (1998b 293) reiterates this when she reports that a translator involved with an MT project eventually "no longer feels threatened by the machine, but has learned to reap as much benefit as possible from what the computer gives him"

Given that only a limited number of methodologies on how to train post-editors have been developed, e g by the EC, Caterpillar, and PAHO, a Special Interest Group (SIG) was set up by the American Machine Translation Association (AMTA) and the European Association for Machine Translation (EAMT) to develop specifications for an optimum post-editing environment, educate audiences who need to know more about post-editing, promote post-editing workshops at conferences, and develop courseware for translation programmes (Allen 2003) [46]

3.6.2 Who Should Be Trained?

Consideration should, of course, be given to the question "who are the target recipients of this teaching?" One might assume that trainee translators should be the primary target audience for post-editing training However, this assumption encompasses an underlying assumption that translation and traditional revision are similar to post-editing and that translators are the best candidates for post-editing It is interesting to consider whether, firstly, translating and post-editing are in fact similar activities and, secondly, whether translator training transfers the necessary skills to an individual for post-editing?

Krings's (2001 360) study on post-editing demonstrates that cognitive processes relating to source-text comprehension during translation and post-editing differ Also, Krings

[45] Personal opinion expressed by members of the Luxembourg based European Commission's Spanish translation department in September 2002

[46] Unfortunately, there has been little activity within this SIG (Jeff Allen, personal communication, November 2002)

concludes that traditional translation is a significantly less linear process than post-editing (ibid 498) Therefore, there is some evidence to suggest that post-editing differs from translation from a cognitive point of view

Post-editing and translation also differ on the practical level Translation usually involves one source text and the creation of one target text to a level of publishable quality Post-editing, on the other hand, involves two source texts, i e the text authored in the source language and the raw MT output, which a translator uses to help produce a final version The task requirements also differ The usual requirement for translation is to produce a target text that meets high quality criteria, whereas post-editing requirements can range from gisting to high-quality publication quality

Translator training focuses on accuracy and equivalence Where specialised translation is the main focus, the trainee translator is taught to be as accurate as possible, where terminology and meaning are concerned, and to aim for cultural and textual equivalence The trainee translator is taught to produce texts suitable for publication It is for this reason that translator training, in the traditional sense, can act as a hindrance to post-editing where the aims are frequently different

Where translation and post-editing do not differ is in the requirement to ascertain the target audience's needs Translation training programmes train translators to examine the expectations of the source language audience and to compare these to the expectations of the target language audience and to translate accordingly Post-editors need to perform this task too

We have seen evidence that post-editing is not the same as translation or traditional revision In fact, some of the demands of post-editing are contrary to the skills and objectives of translators and probably represent one of the reasons why MT implementation has failed in the past Nevertheless, McElhaney and Vasconcellos (ibid 142) believe that there are strong arguments in favour of training translators as post-editors They argue that translators are best able to identify linguistic errors, have a fund of knowledge about the cross-language transfer of concepts, and have the technical resources at their disposal to work efficiently

3.6.3 Skill Sets Required

Krings poses the question does high translation competence indicate high post-editing competence? Drawing on simple concepts from set theory, he puts forward four possible hypotheses concerning connections between translation and post-editing competencies

- They are identical The acquisition of translation competence would automatically lead to a corresponding degree of post-editing competence
- Post-editing competence is a proper subset of translation competence The acquisition of translation competence would lead to a corresponding degree of post-editing competence, but the set would contain additional competence elements that are not needed for post-editing
- Translation competence is a proper subset of post-editing competence The acquisition of translation competence is a necessary subset of post-editing competence, but not sufficient for the latter
- Translation competence and post-editing competence are different but belong to intersecting sets Translation competence is neither a necessary nor a sufficient condition for the acquisition of post-editing competence (Krings 2001 174)

Krings adds that very little research has been done on the connections between personality variables and translation performance even though this is of central importance for predicting a person's suitability as a translator He also mentions the translator's fear of being "reduced to" a post-editor

There is general agreement on the core competencies required for success as a post-editor expert knowledge of the source and target languages and of the subject area (Johnson and Whitelock 1987, Wagner 1987, Olohan 1991) In addition, Olohan (1991) and Vasconcellos (1986a) argue for MT competence as a skill, as well as tolerance Wagner (ibid) reports that translators who are forced to post-edit will not be as efficient as those who have volunteered

Wagner (1987) and Vasconcellos (1986a) both discuss the need for word-processing skills, listing full key proficiency, efficiency in cursor positioning, effective use of search and replace functions and ability to use macros as essential for the skill set of a post-editor In Vasconcellos (1986b), the entire paper is dedicated to the significance of text linguistic knowledge for effective post-editing

There are few differences between the skills mentioned above and those demanded of a professional translator Indeed, Allen and Hogan (2000) have identified a new role in the translation industry which they call "translation post-editor" However, ability to use macros, to code dictionaries for MT, and a positive attitude towards MT are three attributes required of a post-editor that are not usually demanded of a translator

This researcher would argue that several other skills are required for successful post-editing For example, knowledge of MT technology in general would go a long way towards helping the post-editor understand what is going on in the so-called "black-box" and why certain errors occur consistently Understanding the history of MT development, its current status and future prospects would ensure that the post-editor had an appreciation for the technology, its limitations and how it might improve in the future

While most trainee translators are taught the theory and practicalities of terminology management, the trainee post-editor would benefit from an extensive course in machine translation dictionary coding and term base management In any one translation environment, multiple tools can be used to store and retrieve terminology both for source and target text production This presents challenges when terms have to be used across multiple tools and processes which requires knowledge of multiple term management tools and terminology exchange formats, some of which are only emerging at this time (see, for example, the OLIF, TBX, SALT and XLT initiatives) [47]

It has been documented on numerous occasions that authoring source text using controlled language rules improves MT output (Adriaens et al 1996, Mitamura et al 1998, Adrieans et al 2000, Nyberg et al 2003) A drawback to this approach is that authors are unwilling to be constrained by controlled language rules An alternative solution is to use an intermediate editor who has the necessary skills to apply CL rules to a text before it is submitted to MT Being an expert in both source language and target language makes the post-editor a good candidate for this job There is also a significant incentive, i e it reduces the time spent on cleaning up tedious and non-sensical errors in the target language version [48] Therefore, knowledge of controlled languages and controlled authoring tools would benefit post-editors

Vasconcellos (1986a 136) mentions using macros as a necessary skill for post-editors In this researcher's opinion, a post-editor is an ideal candidate for writing macros to automatically clean-up texts since s/he has extensive experience of commonly occurring errors These macros are the first step towards the concept of an automatic post-editing tool, as suggested by Ryan (1988), Knight and Chander (1994), and Allen and Hogan (2000) If equipped with programming skills, the post-editor could develop his or her own program for automatically correcting consistent errors for specific language pairs, text types and MT systems

[47] For information on OLIF see http //www olif net/ (last accessed on October 12 2002) on TBX and SALT see http //www opentag com/tbx htm (last accessed on October 12, 2002), on XLT see http //www ttt org/oscar/xlt/dxlt html (last accessed on October 12, 2002)

[48] Obviously, additional time is required if the post editor is to apply CL rules to the source text The question here is whether it is more or less time consuming and more or less rewarding to apply CL rules than to fix tedious and repetitive MT errors?

As mentioned above, Vasconcellos (1986b) outlines the importance of knowledge of theme and rheme and other language-specific text type norms for post-editing She proposes that post-editing can be performed faster if the information structure of the original sentence is maintained, but this sometimes means that translated elements must take on syntactic roles that are different from their counterparts in the original text She illustrates her point using the source language phrase “Se estudiaran todos los pacientes diagnosticados como ” which would most likely be translated as “All the patients diagnosed as having will be studied” in English However, by changing the syntactic function from a verb to a noun in English (“Studies will be done ”), the information structure of the original can be maintained, post-editing can be completed faster, and the meaning is not compromised (ibid 23) As one can see from this example, a good grounding in text linguistics would be of benefit to post-editors This knowledge could be applied not only in post-editing but also in programming macros and automatic post-editing modules

3.7 CONCLUSION

As more and more organisations automate their translation processes, an increasing number of translators will be asked to post-edit MT output and this will have methodological, professional (vis-a-vis the notions of “quality” and expected throughput) and economic implications (rates paid for post-editing are likely to be lower than those paid for translation)

In Chapter 3 we have given an overview of post-editing research to date, drawing attention to the fact that such research, with the exception of Krings (2001), has focused on defining types of post-editing, categorising MT errors according to post-editing activity, or deriving post-editing “rules” which can be applied manually or with the help of macros Krings’s (2001) research concentrates on the processes involved in post-editing and how they compare with translation processes We agree with Krings’s suggestion that an investigation of post-editing effort should not focus solely on temporal effort, but should also include technical and cognitive effort, and we will address the question of appropriate methodologies for such measurement in Chapter 4

In this Chapter we have also seen evidence that post-editing can produce a TT at a faster rate than human translation (Piggott 1988, Sereda 1982, Wagner 1985) We have also noted claims that MT and post-editing can reduce translation costs (Senez 1998a) As the demand for automation increases, we can expect that more organisations will seek out technology that can produce high quality translation with as little human intervention as possible, as evidenced by the formation of TAUS One way of achieving this is to strive for continued improvement in the performance of TM and MT systems, and much research is being carried out on this, in particular in the domain of MT An additional means of achieving this goal is to improve the input to MT by reducing the number of Negative Translatability Indicators in the ST In doing so, we would expect that the post-editing effort would be lower than for STs where NTIs have not been removed It is this expectation that forms the primary incentive for the current study

Finally, an increase in the demand for post-editing raises questions about the skill-sets and training required for post-editors For example, what is the relationship between translation competence and post-editing competence? What additional skills or traits does a post-editor need? We have also sought to address these questions here

Chapter 4

4. TRANSLATION PROCESS RESEARCH

4.1 INTRODUCTION

The aim of this Chapter is to discuss the relevance of Translation Process Research (TPR) for post-editing research In our Chapter on Post-Editing, we highlighted Krings's assertion that any analysis of post-editing effort should include the parameters of technical, temporal and cognitive effort In essence, Krings suggests that we should investigate *what* the post-editor does, *how* s/he post-edits and *how long* it takes him/her These questions related to translation practice, amongst others, are what translation process researchers have been interested in for some time Given that there are few formal studies of post-editing processes, we turn to TPR to inform our study and, in particular, to consider what methodologies are most appropriate for the study

The background to TPR is first presented and the main topics are discussed in more detail The relevance of these topics for the current study is then considered Think-aloud protocol (TAP), its use as the primary methodology in TPR, and its possibilities and limitations are considered in general and then specifically with regard to the current research We then turn our attention to text production monitoring as a research methodology, describing recent studies that have been carried out using Translog Translog's features are outlined in some detail and we give consideration to its strengths and weaknesses In addition, we describe the methodology known as Choice Network Analysis, which has been proposed as an alternative to TAP Research carried out using Choice Network Analysis is described and we discuss the applicability of this methodology to our current research needs, paying attention again to strengths and weaknesses This leads to a conclusion regarding the research methodologies that are most appropriate for the current study

4.2 BACKGROUND TO TRANSLATION PROCESS RESEARCH

In his book entitled "Translation Performance, Translation Process, and Translation Strategies – A Psycholinguistic Investigation" (1991), Wolfgang Lorscher comments that, as a consequence of translation theory being product- and competence-oriented, hardly any attention had been given to the *process* by which a translation is produced and to translators' actual performance He describes the work carried out in the early and mid-eighties by scholars such as Dechert & Sandrock (1986), Gerloff (1987) and Krings (1986) as "the first approaches to a new type of translation-procedural and performance-analytical research" (Lorscher, ibid 2) In her annotated bibliography, Jaaskelainen (2002) notes that the first research using think-aloud protocol in translation process studies was carried out by Sandrock (1982)

In his earlier review of models of translation, which he criticises as being prescriptive or statically descriptive, Lorscher (1989) again points to the lack of attention given to the translation process In his view, an empirical investigation of the translation process is especially important for three reasons (1) only an empirical investigation of translation performance using a process-analytical approach can allow one to form hypotheses on what is going on in the translator's head, (2) empirical studies into translation have the potential to yield general insights into language processing, and (3) results from this type of research could be used in the teaching of translation

Toury (1991) also speaks in favour of the introduction of empirical studies in translation research "I take the very spread of experimentation as a promise of accelerated movement towards growing non-prescriptivism" (ibid 47) Toury divides empirical translation studies into product-, process- or function-oriented studies which differ in terms of the data they elicit or analyse In product- and function-oriented studies, analysis focuses on *reactions* to translations whereas in process-oriented studies, analysis is applied to the *gradual emergence* of a translated utterance "to the complete neglect of its final version", according to Toury (1991 47)

Toury surveys different methods that can be used for product-, process- and function-oriented empirical studies, including cloze tests, questionnaires and think-aloud protocols A survey of the literature on translation process research by the author of this study reveals that TAP or concurrent "verbalised introspection" is the most favoured research method for gaining insight into the translation process This raises the question as to whether TAP is an appropriate methodology for the analysis of post-editing? We will return to this question at a later stage First, let us present an overview of the central topics in Translation Process Research

4.3 OVERVIEW OF TRANSLATION PROCESS RESEARCH

The topics that are frequently examined under the heading of Translation Process Research are

- Translation Strategies
- Differences between professional and semi-professional translation
- The linearity or non-linearity of the translation process
- The minimax theory

- The measurement of cognitive load
- Text type and length
- The translation assignment
- Data gathering methods

4.3.1 Translation Strategies

Lorscher (1996) identifies a number of translation strategies in his experiment using oral translations from L1 (German) into L2 (English) with subjects who had little training or experience in translating and only partial competence in the TL He defines "translation strategies" as " procedures which the subjects employ in order to solve translation problems" (1991 76ff) Lorscher's experiment is founded on the hypothesis that every individual who has a command of two or more languages also possesses a rudimentary ability to mediate between these languages and that this ability may be considered to be the basis of all translating (ibid 277)

Seguinot (1989b) maintains that the most obvious source of information about translation strategies is a comparison of source and target texts She also believes that monitoring of text production provides an important source of information on the translation process

> *changes made to the text as it is being produced, which includes* [sic] *editing, the crossing out of false starts, and the correction of errors, the rate of production and the timing and duration of interruptions in production, can indicate a change in activity, and sometimes the nature of the activity*
>
> *(ibid 22)*

In an experiment involving only one translator at work and the recording of TAP data, she identifies three "global" strategies The first is a tendency to translate without interruption for as long as possible She saw evidence that the translator starts to produce text when a decision has been made about the first part of the sentence The minimum number of words typed for each sentence before a pause was four

The second is a tendency to correct surface errors immediately, but to leave errors involving meaning until a natural break (unless the error is completely "un-English" as she puts it) The translator concentrates on producing a complete text in one go If errors are detected during text production, the tendency is to correct them only at the end of the line or clause, with the exception of surface errors

The third is a tendency to leave monitoring for qualitative errors to the re-reading stages (ibid 36) Seguinot attributes these strategies to the "principle of least effort", speculating that the correction of superficial errors does not disturb the translation process whereas a critical evaluation of the translation as it is being produced would disturb the process Interestingly for the study at hand, one of her conclusions is that editing is a function of the translation process and that it takes place both during the translation process and upon re-reading

In her report on variation in translation, Seguinot (1997 109) reminds us that translation is a toolbox and that strategies will vary depending on

> *skill, but also on the nature of the assignment, the function of the text, the translating ideology held by the individual or the institution initiating the request, as well as the pragmatics of the translating situation*

Lorscher (1991a), too, addresses variation in strategies In his opinion, "discrete 'when-then' statements [ ] are obviously not possible" Instead we must formulate our generalisations as "When several subjects are faced with a problem X, many or most of them employ similar or the same *types* of strategy" (ibid 280) This is an important comment for a study such as the present one where the number of research subjects, SL and TL pairs and MT engines are necessarily restricted due to limitations on time and scope

Kussmaul in Kussmaul and Tirkkonen-Condit (1995 183) identifies two *types* of activities that can be investigated using TAP, i e surface activities (including the identification of problems, focus of attention, pauses, corrections and the use of a dictionary) and the more elusive hidden activities (macro-planning, solving of problems and cultural transfer)

In his 1987 study, Krings analyses 117 features of translation activity and identifies several different strategies Some of these features are related to the identification, classification and distribution of different problem types Others refer to macro-strategies, i e the organisation of the translation task He also finds evidence of comprehension strategies and decision-making and evaluation strategies He identifies "rephrasing" as the most important strategy for finding translation equivalents and provides "equivalent retrieval diagrams" to illustrate how subjects go about finding equivalents In addition, he identifies a "playing-it-safe" strategy where items suggested by a bi-lingual dictionary with which the subject is unfamiliar are ignored and those with the greatest range of application are selected (ibid 170)

4.3.2 Differences between Professional and Semi-Professional Translation

"Semi-professional" translators are translators who are still undergoing training They are also referred to as "non-professional" or "student" translators in the literature Since most translation process research is carried out in a university environment, it is not surprising that semi-professional translators feature regularly as subjects In some studies, students in the later cycle of translation programmes are treated as "professional" translators (e g Tirkkonen-Condit 1989) The assumption here is that students in their final years of training ought to translate in a manner that is close to that of a professional translator For those who have been successful in recruiting professional translators for process research, data elicited from semi-professional translators provides them with an interesting opportunity for contrast Not only that, but the differences between the strategies used by semi- and professional translators provide translation pedagogues with input for improving their courses for trainee translators Another advantage of using semi-professional translators is pointed out by Jaaskelainen and Tirkkonen-Condit (1991) By comparing the data elicited from both semi-professionals and professionals they were able to identify translation processes that were not evident in the professionals' activities because they had become automatised Seguinot (1997 108) corroborates this finding Jaaskelainen and Tirkkonen-Condit (ibid) note that professionals took account of the task description whereas semi-professionals did not and sometimes struggled with certain aspects of the text as a result

Lorscher (1996) draws a preliminary conclusion that professional and semi-professional translation processes have much in common The differences are to be found in the distribution and frequency of the types of strategies employed by both groups Lorscher's conclusions were that professional translators dealt with larger units of translation than semi-professional translators Students tended not to check TL segments for errors if they had no perceived problem in translating them Professionals, on the other hand did check TL segments He calls this the "ex post realization of translation problems" (ibid 31) and categorises it as an important distinguishing factor for professional and semi-professional translation processes Also, professionals checked TL segments for stylistic and text-type adequacy, whereas semi-professionals restricted their checks to the lexical equivalence and, to an even lesser extent, stylistic levels

Tirkkonen-Condit (1989) sets out to test the hypothesis that the professional and "non-professional" translation processes might differ most in terms of the decision criteria they use Using three types of subjects, professional, semi-professional and non-professional, she finds that professionals take longer to make decisions than the other two types of subject

The professional and semi-professional spend more time on the writing stage than the non-professional The activity of "planning" appears to be relatively automatised for the professional subject This conclusion is based on the fact that evidence for planning appears relatively infrequently in the TAP data for the professional Also, awareness of extralinguistic factors is highest amongst professionals

As we can see, Translation Process researchers have recorded significant differences in the translation strategies of professional and semi-professional translators This has important implications for the current study as there are advantages and disadvantages to using either subject type When considering this question in more detail in Chapter 5, we will draw on the findings from TPR research in order to inform our decisions regarding research methodology

4.3.3 The Linearity of the Translation Process

Seguinot (1997) finds that the translation process is non-linear and iterative Her data suggest that even though a translator may have found a solution to a problem, s/he might continue to look for alternatives and return to that problem This often happens in the middle of another problem-solving episode and not at a natural break She also suggests that translators work on multiple problems at the same time, i e "parallel processing" As already mentioned, Krings found that post-editing is a more linear process than translation (2001 498)

4.3.4 The "Minimax" Strategy

Seguinot (1989b), Krings (1986) and Lorscher (1991a) have all found evidence for the existence of a "minimax" or "principle-of-least-effort" approach to translation Lorscher attributes the apparent dominance of sign-oriented translation in his data to this strategy saying that " subjects generally do not proceed to a deeper level of cognitive processing, which is more abstract and implies a higher cognitive load, before the processing on the higher level has turned out to be unsuccessful or unsatisfactory" (ibid 276) As mentioned previously, Olohan (1991 192) also draws attention to this when she mentions the "Never Mind It'll Do" syndrome

4.3.5 The Measurement of Cognitive Load

TPR has also focused on the topic of cognitive load in translation and pauses and hesitations have been identified as a source of data on cognitive load (Seguinot 1989b, Hansen 2002, Alves 2006) Seguinot speculates that the placement of pauses and hesitations is motivated by natural junctures in the translation process itself and is not necessarily linked to units of meaning For example, a pause might occur in order to cope with a difficulty or because the translator has become aware that something that has already been translated was problematic, another reason might be that the translator has finished a chunk of text and is preparing to move on to another chunk (Seguinot 1989b 32)

Hansen (2002) investigates two hypotheses regarding the occurrence of pauses in the translation process Her first hypothesis (which she subsequently accepts following some empirical analysis) is that some translators demonstrate specific pause behaviour in translation which is independent of language direction Her second hypothesis (which she also accepts) is that there is no correlation between the position, duration and number of pauses and the quality of the translation product

Alves (2006) includes pauses in his analysis of the relationship between cognitive effort, which, following Schilperoord (1996), he terms "cognitive rhythm", and translation product Alves uses a combination of pause analysis and retrospective protocols to analyse translations of one sentence by four translators Although he does not explicitly state it, Alves (ibid 9) assumes that pauses are indicative of cognitive effort

> *The translator showed a pattern of pauses scattered throughout the segment, perhaps an indicator of intense monitoring of the translation process and of issues related to problem-solving and decision making*

Alves (ibid 6) reports that, in a previous study, the cognitive rhythm of novice translators was found to be "erratic" Echoing Hansen's (2002) finding reported above, he observes "that there was no correlation between the subjects' cognitive rhythms and the type of target text rendered by them" (Alves, 2006 6) This finding appears to be confirmed in his 2006 study (ibid 8-9) where one translator produces poor quality output without pausing for a long time while another also produces poor quality output after pausing for 89 seconds

Rothe-Neves (2003) also investigates cognitive resources, or "working memory", and the relationship it might have with performance in translation Rothe-Neves designed a number of tests to investigate whether professional translators have more cognitive resources than novice translators The results indicate that there is no significant relationship between working memory and translation

4.3.5.1 Pauses as Measures of Cognitive Load in Post-Editing

In both spoken and written language production research, it has been claimed that pauses are indicators of cognitive processing (Foulin 1995, Schilperoord 1996, Cenoz 2000) and, as we can see from the short review above, this claim has been adopted by translation process researchers (Seguinot 1989a, Jakobsen 1998, 2003, 2005, Krings 2001, Hansen

2002, Alves 2006) By extension, then, we can assume that pauses are also indicators of cognitive processing in post-editing

Given that little research has been carried out to date on the activity of post-editing, it is not surprising that only one researcher (Krings 2001) comments on the use of pauses in post-editing Krings (2001 304) states that language production research shows that pauses are of great value in the identification of processes, and especially process boundaries In addition, the high operationality of pauses is an advantage for data analysis He uses pauses as markers for identifying "writing acts" in post-editing activity For Krings, the duration of a pause is one second, which, he admits, is an arbitrary unit, but he justifies this by saying that it made sense for his data analysis one second was long enough to identify a distinguishable gap in verbalisation flow and pauses of that length were easy to identify acoustically and to record with relatively reliable intersubjectivity (Krings 2001 210)

Jakobsen (1998, 2005) investigates pauses in the context of translation process analysis using the Translog tool He states that "the assumption that time delay during text production and translation correlates with cognitive processing is strongly supported by the systematic syntagmatic distribution of delays" (ibid 100) In his article describing how Translog records pauses, he claims that a pause unit of 0 20 seconds brings us close to many subjects' typing speed He also suggests that a pause length of 1 second is appropriate for observing delays in a text production event

> *For the purpose of observing the distribution of longer delays in a text production event, a representation with a 1 second time unit will often turn out to be very appropriate because it represents all the delays we want to identify and suppresses most of the delays we are not interested in*
>
> *(ibid 83)*

On the other end of the scale, Jakobsen argues that a time delay greater than 10 seconds will identify text initial and text final delays, delays between paragraphs, and delays appearing less systematically in front of particularly difficult text segments (ibid 84) In his 2003 article, Jakobsen investigates how time is divided between translation phases across semi-professional (senior cycle students) and professional translators The phases he identifies are the initial orientation phase, the middle drafting phase and the final revision phase Jakobsen observes a difference between the professional and semi-professional translators in the allocation of time between the three phases He reports that, on average, professional translators dedicated more time to the initial phase and less time to the drafting phase than semi-professional translators

Jakobsen (2005) uses pauses (recorded by Translog) as segment delimiters in his analysis of the "peak performance" (ibid 114) of expert translators, which he compares with non-expert (i e student) translators Segments are measured according to the number of text production keys pressed and segments containing 60 or more text production keystrokes are counted as "instances of peak performance" (ibid 113), otherwise described by Jakobsen as "the expert translator's ability to suddenly sprint into extended creative performance" (ibid 114) Instances of peak performance are interesting, according to Jakobsen, because they are not as prevalent in non-expert translators' performance and, therefore, offer an insight into the differences between expert and non-expert behaviour

While it is beyond the scope of this dissertation to provide an extensive overview of research on pauses in the domain of language production, it is perhaps worthwhile mentioning some research in that domain, as more extensive research has been carried out in language production than in the domains of translation or post-editing, and we can use the general findings to inform our observations on pause activity in post-editing

Regardless of whether one is more interested in oral language production or written language production, there appears to be some agreement that pauses are influenced by a number of different factors For example, Foulin (1995) reports on other studies that have demonstrated that pauses and hesitations increase as the subject matter complexity increases The number of pauses also seems to increase as familiarity with the subject matter decreases or as the context becomes more abstract The same tendencies are noted in both written and oral production

Foulin (ibid) points to the relative lack of interest in the revision process in studies on written language production He attributes this to the dominant idea that pauses reflect conceptual and linguistic forward planning, but points to some research (Kaufer et al 1986) in which systematic relations between pauses and back revision were revealed Foulin (ibid 494) suggests that an analysis of pauses needs to take into consideration whether the pause has to do with forward planning or with revision

> *En somme, des variations de la durée des pauses, principalement celles localisees en fin d'unites structurales, pourraient egalement survenir en fonction de ce qui precede Des lors, il conviendrait, d'une part de decrire et expliquer le fonctionnement de ces pauses retrospectives et, d'autre part, de determiner dans quelle mesure l'activite de pause en un site donne depend de decisions relevant de la planification du texte subsequent ou du contrôle du texte antecedent* [49]

[49] In summary, variations in the duration of pauses, especially those located at the end of a structural unit, could arise from what went before Therefore, one ought to describe and explain the functions of these pauses on the basis of what preceded them and in addition determine to what extent pause activity in a particular location is a function of forward planning or text revision [my translation]

Van Waes and Schellens (2003) compared revision activity between pen and paper writers and computer writers and found that computer writers tend to revise more at the level of the letter than pen and paper writers Also, large numbers of short pauses occurring in rapid succession within sentences were recorded for computer writers This leads Van Waes and Schellens to conclude that computer writers use a writing process in which planning, formulation and revision are strongly focused on relatively small units of text

Although it is generally agreed that pauses are indicators of cognitive processing, and should, therefore, be included as a parameter in the research question addressed in this dissertation, it is hopefully clear from this short review of literature on the subject that pauses can be influenced by a number of parameters and that current methodologies do not allow us to specify exactly what motivates a particular pause [50] We will return to the subject of pauses in Chapter 6 (Data Analysis)

4.3.6 Number and Type of Subjects

The number of subjects used on average is deemed to be small when compared to research group sizes in the field of psychology, for example (Krings, 2001 70) This is one of the unavoidable shortcomings of using think-aloud protocol in translation process research

> *The reason for the small number of subjects is clear the Think-Aloud Protocol is a data acquisition method that involves an unusually great amount of effort The transcription of the verbalizations makes it almost impossible for the individual researcher to work through a sufficiently large subject group in an acceptable amount of time This shortcoming naturally has severe consequences for the generalizability of any results*
>
> *(Krings ibid)*

Krings used 52 subjects altogether in his study However, it should be emphasized that the primary research group was composed of only 16 people (ibid 71) Krings also addresses the issue of the general lack of availability of professional translators He insists that there are valid arguments *against* the use of professional translators in translation process research (TPR) Firstly, there is the psychological barrier to research As Krings puts it

> *Like teachers who dislike being observed once they have completed student teaching, many professional translators don't care to have someone observe them at their craft This is understandable They are, after all, granting an outsider a detailed look at their professional technique, with all the strengths, but also with all the weaknesses and shortcomings that it possesses*
>
> *(ibid 72)*

[50] For a more detailed discussion on pauses as indicators of cognitive effort in post editing effort, see (O Brien 2006)

Secondly, a group of professional translators represents a decidedly less homogenous group when compared with a group of students in the same course and at the same level Results from the professional translators would, therefore, be less generalizable than results from a more homogenous student group

A third argument sometimes proposed against the use of professional translators is that members of this group would display increased automaticity in the translation process and automatic processes are not available in Short-Term Memory (STM) for verbalisation In her research, however, Gerloff (1988) determined that this assumption was false

4.3.7 Text Type and Length

The texts used in many of the studies surveyed by Krings were from newspapers, journals, tourist information or linguistic monographs Most studies used one text only, some used two and the total number of words ranged from 60-750 Some of Krings's texts had a word count greater than 1,000, but not all words were translated during the experiments Text type selection and word count are two criteria that will be taken into account in Chapter 5 when we are considering the design of this study

4.3.8 Translation Assignment

Studies by Jaaskelainen (1987) and Tirkkonen-Condit (1989) demonstrated that the translation assignment, i e instructions to the subjects on what the translation was *for*, strongly influences the translation process Yet many of the studies surveyed by Krings did not provide the subjects with a translation assignment Krings argues that a "meaningful translation assignment should be a standard component of translation process research design" (ibid 75) In our Chapter on Methodology (Chapter 5), we will take into account the importance of presenting a translation assignment to the subjects

4.3.9 Data gathering methods

Most of the studies presented by Krings and mentioned above in the Overview of Translation Process Research use TAP as the primary data collection method Over half of the studies presented by Krings supplemented the TAP data with data derived from other methods, e g interviews, questionnaires, dialogue protocols (verbal protocols produced by pairs of subjects – House 1988) More recently, the trend in TPR has moved towards triangulating TAP data (concurrent and retrospective verbalisations) with keyboard logging data and product analysis (Hansen 1999a, 1999b, 2002, 2003, Jakobsen 1998, 1999, 2003, 2005a, 2005b, Alves and Gonçalves 2003, Alves 2006)

TAP requires subjects to verbalise their thoughts about the action that they are performing at a particular moment in time The use of introspection via think-aloud protocols to gain access to a subject's mind dates from the late 1800s Around the 1940s, however, behaviourism took over as the epistemological and methodological paradigm in psychology and introspection was not used again until half a century later (Krings 2001 215)

Translation Process researchers have been drawn to the use of TAP as a method of gaining insight into translation strategies since the early 1980s In this type of study, the subject is asked to verbalise the thoughts going through his/her mind during the translation process These verbalisations are recorded and transcribed by the researcher The subject's focus of attention, e g source text, target text, dictionary etc , may also be video-taped An analysis of the resulting data is carried out and the results are often correlated with the static text that was produced by the end of the translation process

Jaaskelainen and Tirkkonen-Condit (1991) report that the methodological framework most often used for the elicitation of TAP (or "verbal report data") is based on Ericsson and Simon's information processing model of 1984 which describes how information is stored and retrieved from both short-term memory (STM) and long-term memory (LTM) Ericsson and Simon (ibid 16) contend that undirected, concurrent verbal reports provide the most direct access to human thought processes They further contend that, for a thought to be verbalisable, it must be contained in STM However, this can be prevented from happening if the cognitive process has been "automated", and this phenomenon forms one of the main arguments against the use of TAP We will discuss the arguments for and against the use of TAP in more detail later First, however, let us consider the relevance of TPR research to this study

4.4 Possibilities and Limitations of Verbal Data

Given the dominance of TAP as a method for collating data in translation process research, some consideration must be given to the possibilities and limitations of this method Krings states that when using introspective methods it is important to provide justification because the validity of these methods has been questioned in the past

The first advantage associated with the use of Verbal Report Data (VRD) is that they provide a large volume of information on cognitive processes Krings points out that another argument in favour of the use of VRD is that an analysis of the product of language processing alone does not provide any insight into the processes that went towards *forming* that product Also, some processes result in no physical product Krings (2001 217) puts it like this

> *While mere observation of behaviour only records and analyzes the physical products of these processes, verbal-report data also offers an insight into the creation of these products Verbal-report data provides the motion picture to a photograph of a scene, so to speak, showing how the scene came to be*

The third advantage listed is the unstructured nature of TAPs, which some would deem initially to be a disadvantage Krings argues that this method of data collection predetermines the researcher's model far less than other methods TAP is therefore well suited towards hypothesis-generating studies In addition, verbal-report data can also give the researcher access to the "subjective theories" of the subjects involved

The validity of VRD has been questioned by its opponents Opponents of VRD insist that there is no systematic connection between verbalisation and mental processes or action-determining cognition How can we tell that what a subject says during an experiment using verbal-report techniques is actually what that subject is thinking or doing at the time of utterance?

Krings breaks the problems regarding VRD into three areas

1 The problem of consistency or epiphenomenality Do consistent (systematic) connections exist between verbalisations and cognitive processes?

2 The problem of interference Does performance of the verbalisation task itself change the normal course of cognitive processes?

3 The problem of completeness Can verbalisations contain complete records of cognitive processes?

Drawing on the work of Ericsson and Simon (1984) as his theoretical and methodological basis, Krings addresses each of these points in turn On the subject of consistency, Ericsson and Simon found that VRD are not responsible for inconsistencies Rather, it is the way in which these data are used that is problematic Their studies revealed that experimental findings used to argue against verbal-report data actually refer to retrospective verbalisation and not to simultaneous thinking-aloud Also, data validity can be compromised by asking subjects to verbalise processes that are automatic under normal conditions This can result in the subject being forced to construct a plausible description In summary, Ericsson and Simon's central statement is that the highest level of validity can be expected when data are collected as close to the primary task as possible, if the data are not abstract or general, and if they only refer to consciously occurring processes Ericsson and Simon also assume that data validity is especially high if no re-coding is involved, e g the subject does not have to re-code a visual code into a linguistic code before describing it Krings's argument is that, based on Ericsson and Simon's findings, thinking-aloud is the process that best fulfils the requirements for a high level of validity "In conclusion, it can be said that the use of thinking-aloud with language-related cognitive processes is the form of verbal-report data use having the least problems with data validity" (Krings, 2001 226)

The problem of interference is also addressed by Ericsson and Simon who counter the claim that thinking-aloud is an unnatural task by reporting that they frequently observed quiet but audible verbalisation in their tests when no verbalisation was requested Krings's observations in his own study back this up The fact that subvocalisation with lip and larynx movements is very common in text reception activities is also mentioned to counter the claim that thinking-aloud is an unnatural task

It is acknowledged that interference does occur to the extent that there is a deceleration of the primary process, known as the "slow-down effect" In Krings's experiments, it was noted that the group using thinking-aloud needed one third more time to process a task than the group not using thinking-aloud This increase in time is not seen as a problem by Krings, but he does ask the question whether or not there is also a qualitative interference as a result of using thinking-aloud? The answer proposed is that there is no qualitative interference since studies show no differences between the quality of solutions implemented by groups using thinking-aloud and by those not using thinking-aloud However, Krings acknowledges the fact that some quantitative interference does occur his study showed that those using TAP during post-editing produced twice as many revisions than those not using TAP

The question of completeness of records has not yet, in Krings's view, been sufficiently resolved It is true that different subjects will verbalise to greater or lesser extents, but does this mean that those who verbalise less have incomplete verbalisations? Krings's study showed that those who verbalised to a greater extent than others, did so consistently across different texts This suggests that some subjects are more willing or able to verbalise than others Whether or not the willingness to verbalise is linked to personality was addressed by a study quoted in Krings (Gilhooley 1987) Gilhooley did not find any connection between verbalisation volume and personality characteristics Krings concludes that indications show that think-aloud data on cognitive processes occurring in short-term memory are not always complete and this must be taken into account during data analysis

By observing processing speeds for post-editing and relating them to the use or non-use of think-aloud protocols in his own study, Krings establishes that the use of TAP can considerably influence processing time, increasing it by 31% (known as the "slow-down effect") (Krings, 2001 279) However, it was also found that the slow-down effect decreased from one text to the next leading Krings to comment that

> *This effect allows for the hypothesis that Thinking Aloud causes a habituation effect, and/or that unwillingness to verbalize decreases during the text period and the required time for a task with Thinking Aloud approaches that of the same task without Thinking Aloud.*
>
> *(ibid: 279)*

Another important finding reported by Krings that is relevant to the current study is that all important forms of non-linear writing (deletion, insertion and overwriting) occurred significantly more frequently when subjects used TAP than when they did not. This leads to the question: is there then a direct influence from TAP on revision behaviour? If there is, this would obviously have an impact on the data gathered for the present study. Krings also observes that the number of elemental writing acts required per hundred words of target text is significantly lower when subjects do not use TAP. This leads Krings to hypothesize that "Thinking Aloud not only externalizes cognitive processes during text production, but at least partially also changes them" (Krings, 2001: 527).

Bernardini (2001) criticises what she sees as a *methodological* failing with regard to the use of TAP in translation studies. She remarks that TAPs are only considered to be valid if they are collected under very rigorous experimental conditions, but the mode of recording TAPs in translation studies tends to be "relaxed" (ibid: 252), something which may result in the invalidation of the results obtained.

Krings concludes that there are specific possibilities and limitations associated with VRD, as with all other research methods. Until some of the concerns regarding the validity of thinking-aloud have been further allayed, Krings promotes triangulation whereby different methods of assessment are combined in order to validate one set of findings against another.

4.4.1 TAP as an Appropriate Methodology for the Current Study

Clearly, TAP has been used in TPR to significant advantage and has revealed some interesting results regarding what takes place in a translator's mind during the translation process. Nevertheless, there are a number of drawbacks associated with this technique, many of which have been touched on above. Some of these drawbacks are more relevant for the study at hand than others and shall be addressed here.

On a general level, the ability to verbalise detailed and relevant protocols has been found to be largely personality-specific. Therefore, one may get different results depending on the subjects recruited for the study.

House (1988) has highlighted the artificiality of the TAP-producing situation While a translator may have an internal dialogue going on in his or her head during translation, this is seldom verbalised The importance of creating a realistic working environment for the current study is addressed in Chapter 5 Introducing any artificiality would be contrary to this objective

Toury's (1991) contention that interference can be caused by the two different modes (i e written vs spoken communication) involved in translating and verbalising is certainly something to be considered Not only that, but Ericsson and Simon (1984) have admitted that even with Level 1 verbalisation (i e verbalisation involving the articulation of information stored in verbal mode), one might expect a decrease in task performance Most importantly, Krings found that the use of TAP interfered both with the temporal effort of post-editing and with the technical effort in that all important forms of non-linear writing (deletion, insertion and overwriting) occurred significantly more frequently when using TAP than when not Since the focus of the current research is the correlation between the effort involved in post-editing MT output and source text translatability, and two of the ways in which such effort is measured include cognitive load and the time required for task completion, it is desirable to eliminate factors that are likely to add to post-editing effort

Some of the aforementioned problems with the use of concurrent TAPs could be redressed by using retrospective TAPS, which is the method preferred by several TP researchers, in particular those who triangulate results using Translog (discussed in more detail later) and TAPs (e g Hansen 1999b, Jakobsen 2003, Alves 2003, 2006) However, retrospective protocols still present problems from a methodological point of view because they cannot compensate for individual levels of verbalisation willingness or for the inability of some subjects to recognise that they had a problem and to verbalise their solutions to it Also, it is claimed that subjects cannot verbalise all thoughts retrospectively because some thoughts would have moved into long-term memory by the time the subject is asked to verbalise (Bernardini 2001) An additional problem is the length of time required to transcribe and analyse protocols in general

When these findings are added to the other reservations on TAP as a methodology, it appears that alternatives to both concurrent and retrospective TAPs should be explored

4.5 TEXT PRODUCTION MONITORING AS A RESEARCH METHODOLOGY

In general terms, text production monitoring can be understood as the capture and analysis of data on how a text is produced by a subject, including pauses in production, hesitation, deletions, rewriting of text etc. One tool which has been developed for monitoring text production in general, and translation more specifically, is Translog.

Jakobsen (1999) discusses the development and use of Translog as an aid to translation process research. Translog is a tool that was designed specifically to record the translation process as it evolves in a computerised text production environment. One of the motivating factors behind its development was the need for quantitative reinforcement of assumptions about translating which were based on qualitative data only (ibid: 11). With the development of Translog, Jakobsen was aiming for a dialogue between qualitative and quantitative approaches that would result in a "synergetic refinement" of both methods (ibid: 11).

One of the functions of Translog is to record the time required to complete a translation task. The underlying assumption behind this time recording is:

> *...(a) that there is a general correlation between time delay during text production and the cognitive processing that is involved and (b) that – more specifically – this correlation can be observed at different levels.*
>
> *(ibid: 14)*

Jakobsen outlines three stages in the data collection process using Translog. Stage one involves preliminary testing. Subjects are asked to write a piece of text using the program so that data on their typing speeds, editing and general navigation skills can be collected. Stage two involves logging data regarding the translation process itself. This stage can be combined with other methods of simultaneous data collection such as TAP, audio or video recordings and direct observation. Stage three involves a retrospective interview with or without a replay of the data logged by Translog. Immediate retrospection reinforced by observation of a replay of the subject's own text production process results in rich data, according to Jakobsen.

Jakobsen (2003) uses Translog to investigate the effects of thinking-aloud on translation speed, revision and segmentation. In keeping with Krings's findings, he reports that thinking-aloud slows down TT production for semi-professionals and experts alike and that it also results in an increase in the number of segments in text production – suggesting that smaller chunks or units are dealt with when subjects have to think aloud. Interestingly, he finds that thinking-aloud has no significant effect on the revision process.

Hansen (1999b) describes an experiment where Translog was used in conjunction with retrospective techniques to see if these combined methods were useful for observing macro- and micro-translation strategies. She concludes that the combination of the two methods was useful in providing insight into the degree of consciousness a translator might have about a translation problem

Hansen (2003) reports on an experiment where Translog was used along with retrospective commentaries in order to investigate individual and general competence patterns She correlated this data with information on subjects' backgrounds Hansen notes that where her subjects had problems that caused long pauses, the subjects were unable to express what the problem was during retrospection She argues that the translator needs to be aware of the problems encountered and to have the necessary vocabulary/metalanguage to be able to describe those problems On the use of Translog in conjunction with other methods, Hansen states (ibid 27)

> *The possibility of combining introspective methods, TAPs and retrospection, with a computer program like Translog , has changed and improved the study of translation processes The computer software provides us with quantitative, more objective data about processes, allowing us with its "view function" to see all movements, corrections or changes as well as the position and length of all phases and pauses during the process*

Alves and Gonçalves (2003) investigate the characteristics of problem solving and decision making in translation using Translog and retrospective protocols By examining both the retrospective protocols and the Translog log files of four novice translators they show that it is possible to map the recursive movements of translators and to identify parameters of relevance in their problem solving and decision making processes (ibid 21) As mentioned earlier, Alves (2006) uses Translog to investigate the relationship between the cognitive rhythm of translators and the type of target text rendered by them

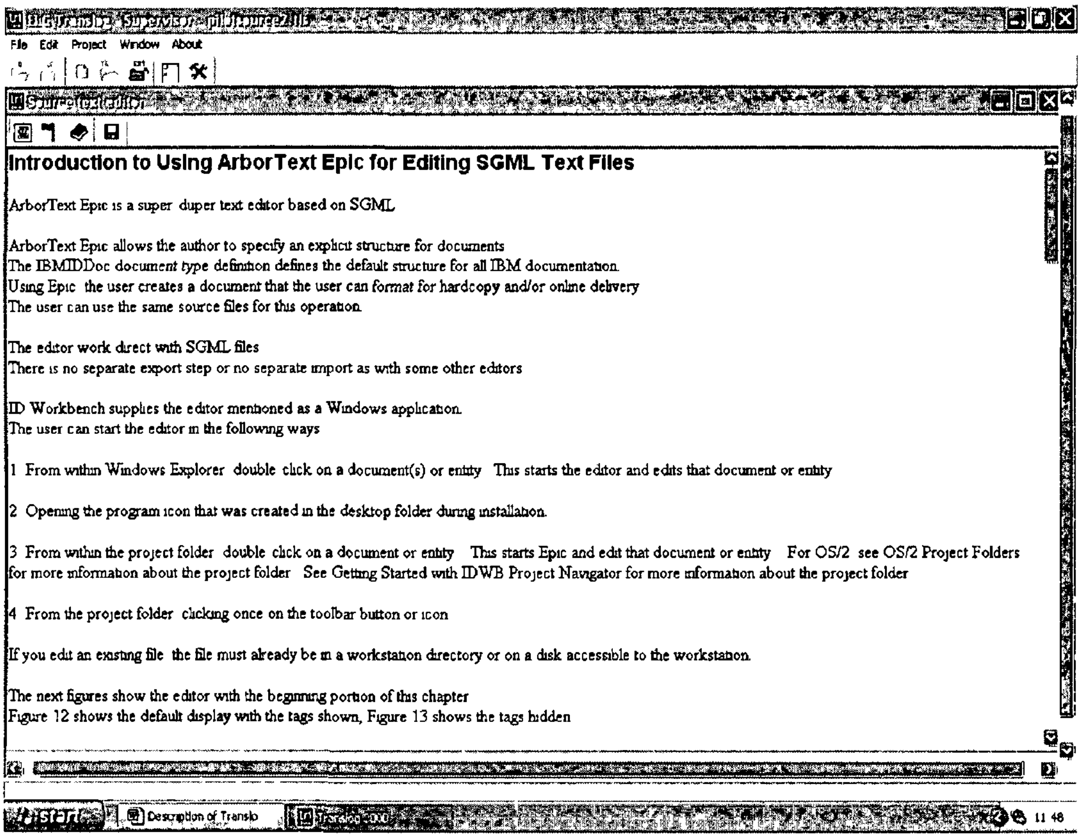
4.5.1 Description of Translog

Translog is made up of two applications *Translog User* and *Translog Supervisor* In general, Translog Supervisor is used to set up projects while Translog User is used by translators as a text editing environment, and to record the translators' text processing activity during translation

4.5.1.1 TRANSLOG SUPERVISOR

As indicated above, Translog Supervisor is used to set up projects In practice this means selecting a source text, entering source and target terms in a dictionary, and specifying how translation segments should appear to the user and what units of time should be used to record translation activity

The first step in setting up a project involves the addition of a source text to the project definition This can be done either by writing the text in the source text window or by pasting the text from another application, e g MS Word, into the source window Figure 4 1 below shows the source document used for the current research pasted into the Translog Supervisor source text window



Introduction to Using ArborText Epic for Editing SGML Text Files

ArborText Epic is a super duper text editor based on SGML

ArborText Epic allows the author to specify an explicit structure for documents
The IBMIDDoc document *type* definition defines the default structure for all IBM documentation.
Using Epic the user creates a document that the user can *format* for hardcopy and/or online delivery
The user can use the same source files for this operation.

The editor work direct with SGML files
There is no separate export step or no separate import as with some other editors

ID Workbench supplies the editor mentioned as a Windows application.
The user can start the editor in the following ways

1 From within Windows Explorer double click on a document(s) or entity This starts the editor and edits that document or entity

2 Opening the program icon that was created in the desktop folder during installation.

3 From within the project folder double click on a document or entity This starts Epic and edit that document or entity For OS/2 see OS/2 Project Folders for more information about the project folder See Getting Started with IDWB Project Navigator for more information about the project folder

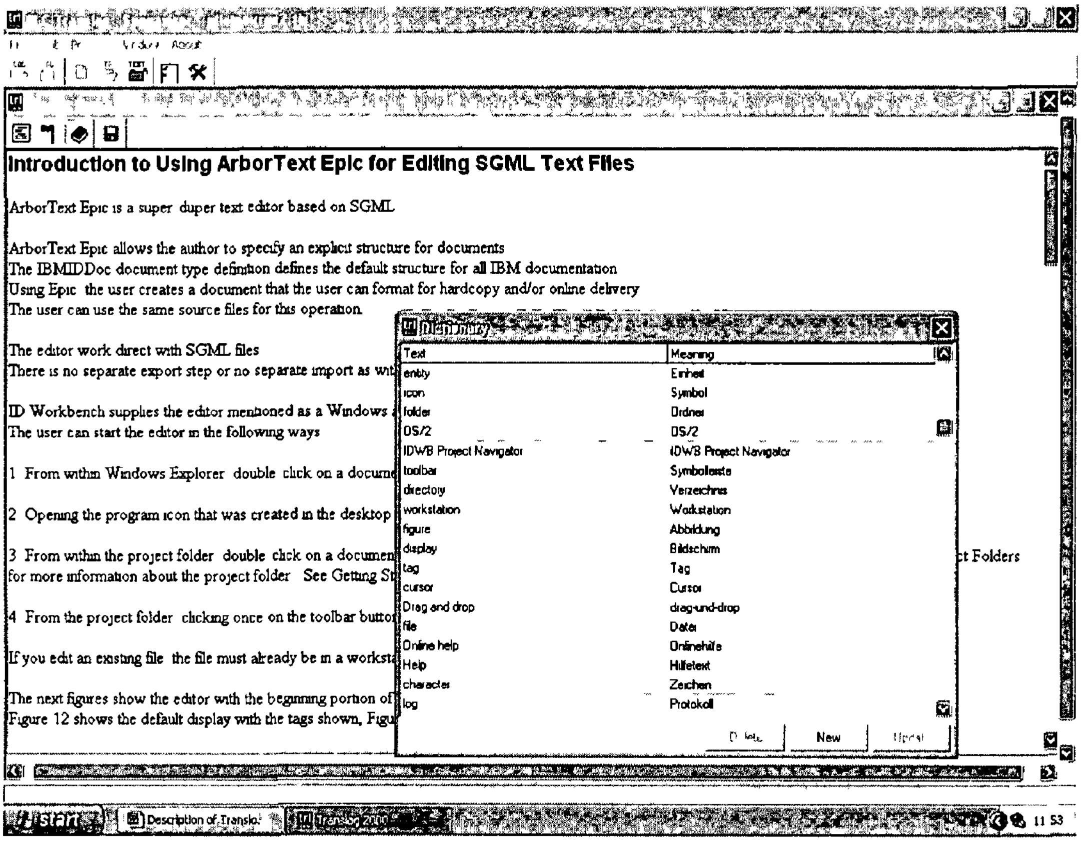
4 From the project folder clicking once on the toolbar button or icon

If you edit an existing file the file must already be in a workstation directory or on a disk accessible to the workstation.

The next figures show the editor with the beginning portion of this chapter
Figure 12 shows the default display with the tags shown, Figure 13 shows the tags hidden

Figure 4 1 Source text in Translog Supervisor

When setting up the source text, the user can use a red flag button (top left toolbar) to set his or her own translation segment markers The dictionary icon allows one to input source and target terms in the dictionary using a rudimentary interface (see Figure 4 2)



Introduction to Using ArborText Epic for Editing SGML Text Files

ArborText Epic is a super duper text editor based on SGML

ArborText Epic allows the author to specify an explicit structure for documents
The IBMIDDoc document type definition defines the default structure for all IBM documentation
Using Epic the user creates a document that the user can format for hardcopy and/or online delivery
The user can use the same source files for this operation

The editor work direct with SGML files
There is no separate export step or no separate import as wit

ID Workbench supplies the editor mentioned as a Windows
The user can start the editor in the following ways

1 From within Windows Explorer double click on a docume

2 Opening the program icon that was created in the desktop

3 From within the project folder double click on a documen ... t Folders
for more information about the project folder See Getting St

4 From the project folder clicking once on the toolbar butto

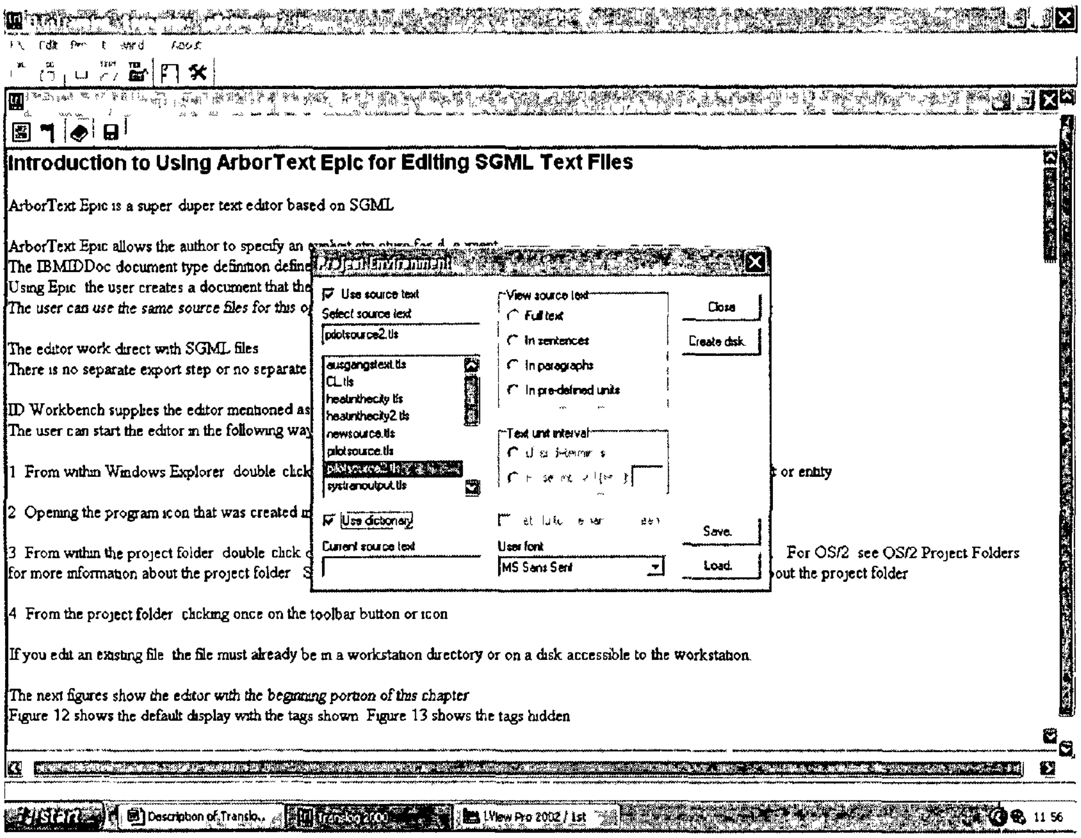
If you edit an existing file the file must already be in a workst

The next figures show the editor with the beginning portion of
Figure 12 shows the default display with the tags shown, Figu

Text	Meaning
entity	Einheit
icon	Symbol
folder	Ordner
OS/2	OS/2
IDWB Project Navigator	IDWB Project Navigator
toolbar	Symbolleiste
directory	Verzeichnis
workstation	Workstation
figure	Abbildung
display	Bildschirm
tag	Tag
cursor	Cursor
Drag and drop	drag-und-drop
file	Datei
Online help	Onlinehilfe
Help	Hilfetext
character	Zeichen
log	Protokoll

Figure 4 2 Adding terms to the dictionary in Translog Supervisor

When the source text and dictionary have been prepared, they are packaged together into a "Translog Project" using the Project Environment Dialog Box (see Figure 4 3)



Introduction to Using ArborText Epic for Editing SGML Text Files

ArborText Epic is a super duper text editor based on SGML

ArborText Epic allows the author to specify an
The IBMIDDoc document type definition define
Using Epic the user creates a document that the
The user can *use the same source files for this o*

The editor work direct with SGML files
There is no separate export step or no separate

ID Workbench supplies the editor mentioned as
The user can start the editor in the following way

1 From within Windows Explorer double click … or entity

2 Opening the program icon that was created in

3 From within the project folder double click … For OS/2 see OS/2 Project Folders
for more information about the project folder S … out the project folder

4 From the project folder clicking once on the toolbar button or icon

If you edit an existing file the file must already be in a workstation directory or on a disk accessible to the workstation

The next figures show *the* editor *with the beginning portion of this chapter*
Figure 12 shows the default display with the tags shown Figure 13 shows the tags hidden

☑ Use source text
Select source text
pilotsource2.tls

ausgangstext.tls
CL.tls
healthcity.tls
healthcity2.tls
newsource.tls
pilotsource.tls
pilotsource2.tls
systranoutput.tls

View source text
○ Full text
○ In sentences
○ In paragraphs
○ In pre-defined units

Text unit interval

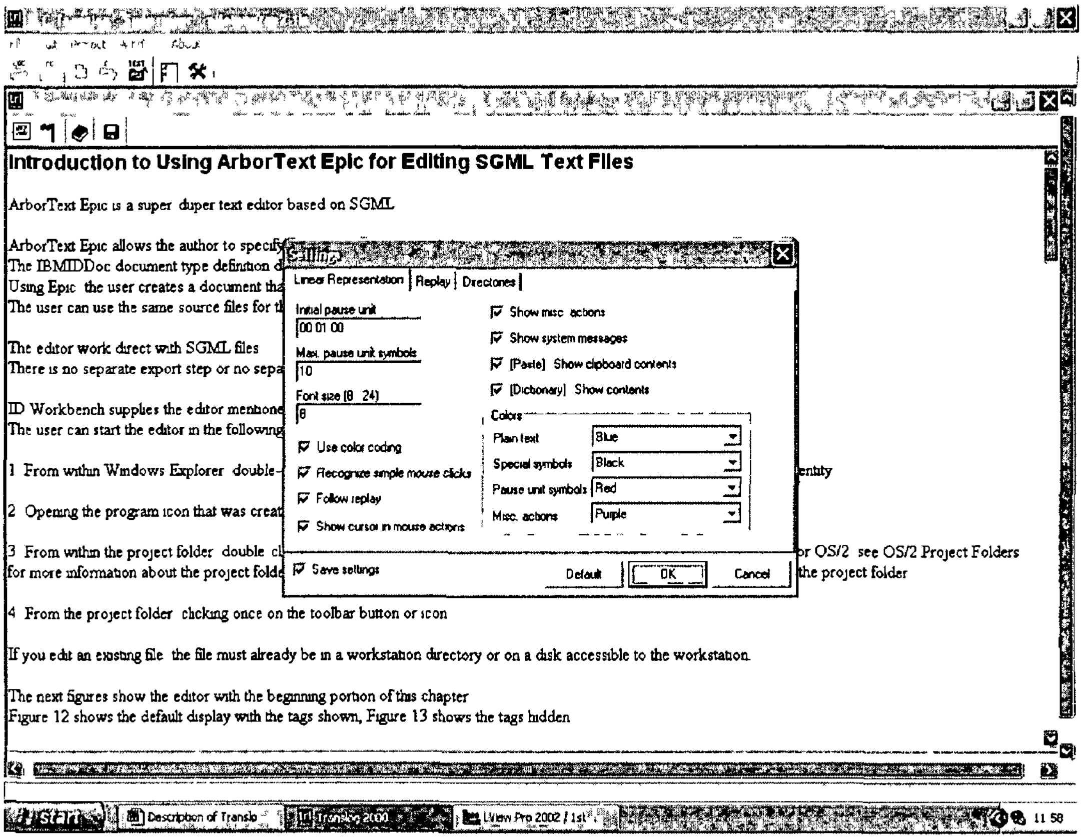
☑ Use dictionary
Current source text

User font
MS Sans Serif

Figure 4 3 The Project Environment Dialog Box

This dialog box allows the user to specify what source text is to be used and whether or not the dictionary should be incorporated into the project The user also specifies whether the source text should be presented to the translator as full text, sentence by sentence, paragraph by paragraph, or in pre-defined units (using the translation unit markers mentioned previously) One can also determine the timing of the display of text – either the user manually determines when the next segment of text should be displayed or the units can be displayed at pre-determined intervals

The global settings for logging translation activity can also be set using Translog Supervisor (see Figure 4 4)



Introduction to Using ArborText Epic for Editing SGML Text Files

ArborText Epic is a super duper text editor based on SGML

ArborText Epic allows the author to specify
The IBMIDDoc document type definition d
Using Epic the user creates a document tha
The user can use the same source files for th

The editor work direct with SGML files
There is no separate export step or no sepa

ID Workbench supplies the editor mentione
The user can start the editor in the following

1 From within Windows Explorer double- … entity

2 Opening the program icon that was creat

3 From within the project folder double cl … or OS/2 see OS/2 Project Folders
for more information about the project folde … the project folder

4 From the project folder clicking once on the toolbar button or icon

If you edit an existing file the file must already be in a workstation directory or on a disk accessible to the workstation.

The next figures show the editor with the beginning portion of this chapter
Figure 12 shows the default display with the tags shown, Figure 13 shows the tags hidden

Settings

Linear Representation | Replay | Directories

Initial pause unit: 00 01 00
Max. pause unit symbols: 10
Font size (8 24): 8

☑ Use color coding
☑ Recognize simple mouse clicks
☑ Follow replay
☑ Show cursor in mouse actions

☑ Show misc actions
☑ Show system messages
☑ [Paste] Show clipboard contents
☑ [Dictionary] Show contents

Colors
Plain text: Blue
Special symbols: Black
Pause unit symbols: Red
Misc. actions: Purple

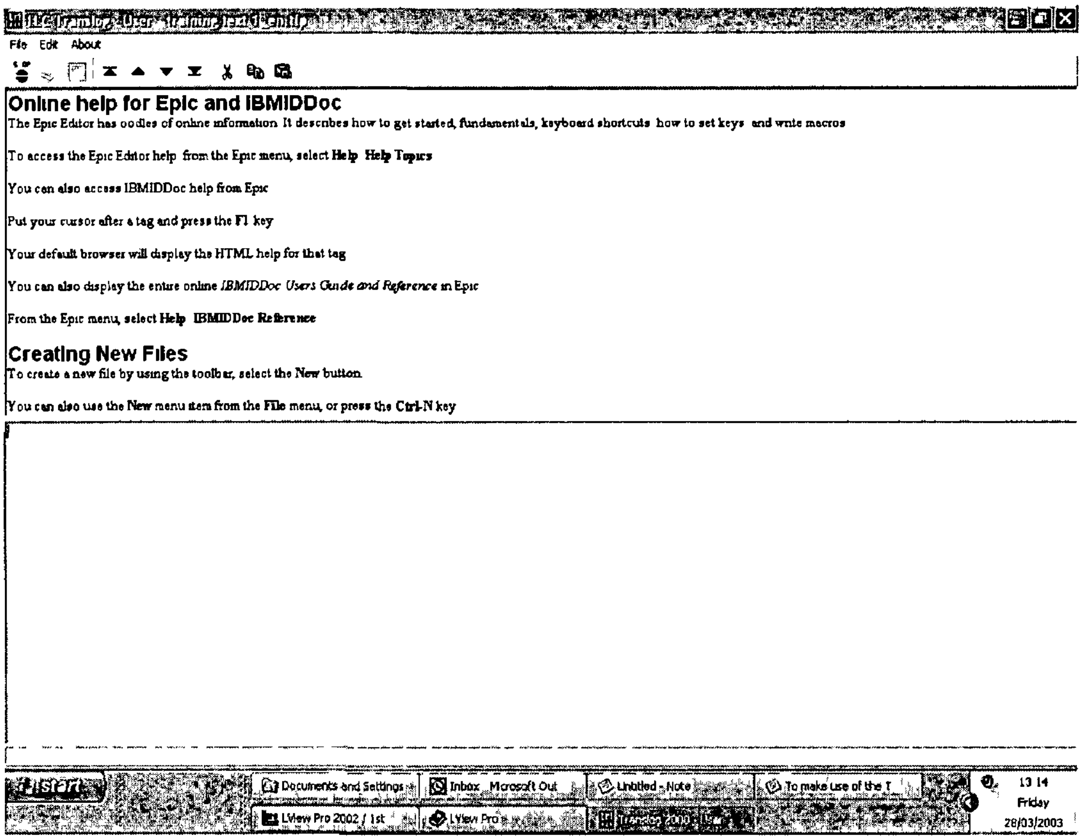
☑ Save settings

Figure 4 4 Translog Linear Representation Settings

Using the Settings Dialog Box, the user can instruct Translog to log system messages, the contents of pasting actions, the contents of dictionary look-ups etc

4.5.1.2 TRANSLOG USER

Once a project has been set up in Translog Supervisor, it is then opened in Translog User and the source text is translated The Translog User interface is rudimentary (see Figure 4 5)



Online help for Epic and IBMIDDoc

The Epic Editor has oodles of online information It describes how to get started, fundamentals, keyboard shortcuts how to set keys and write macros

To access the Epic Editor help from the Epic menu, select **Help Help Topics**

You can also access IBMIDDoc help from Epic

Put your cursor after a tag and press the F1 key

Your default browser will display the HTML help for that tag

You can also display the entire online *IBMIDDoc Users Guide and Reference* in Epic

From the Epic menu, select **Help IBMIDDoc Reference**

Creating New Files

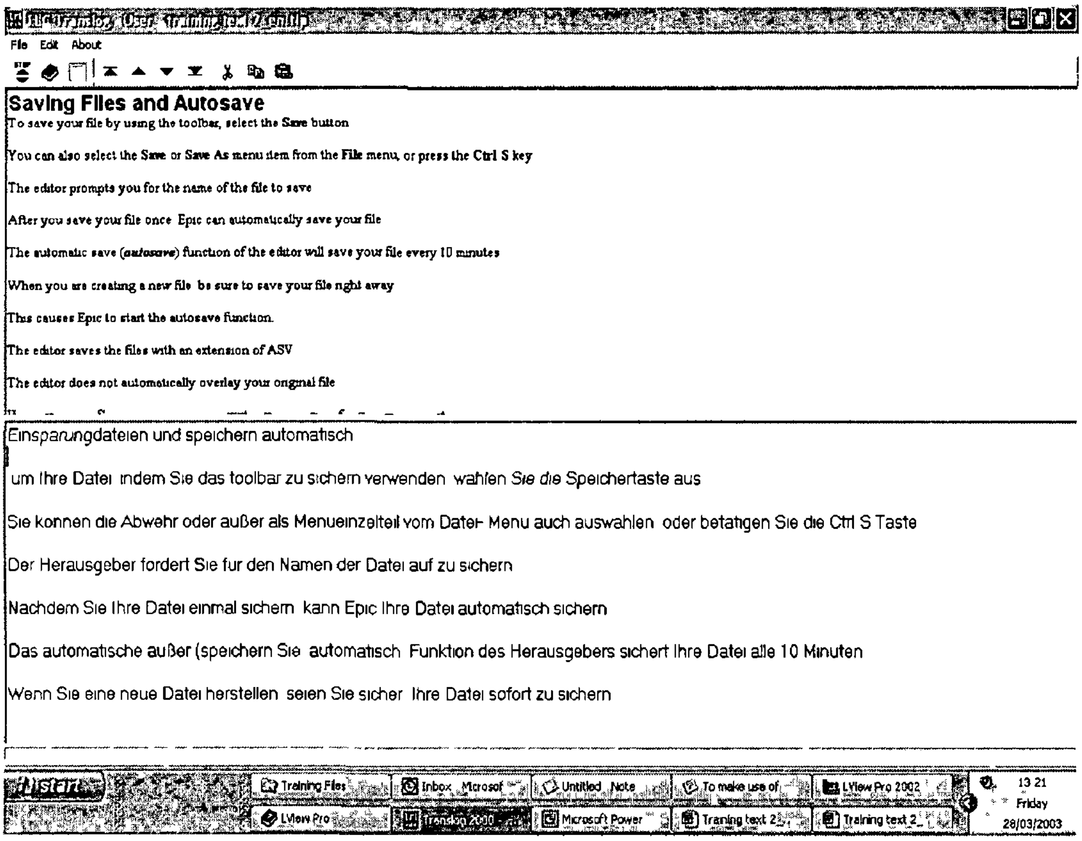
To create a new file by using the toolbar, select the New button

You can also use the New menu item from the File menu, or press the Ctrl-N key

Figure 4 5 The Translog User Interface

When the translator is ready, the green flag on the toolbar is clicked and this prompts Translog User to commence the logging process While working, the translator can write, delete, move, copy, cut and paste text S/he can also look a term up in the dictionary or scroll through the source and target texts When the translation is complete, the “Stop” sign is clicked and Translog User prompts the user to save the log file

For the monitoring of post-editing activity, the only difference in the use of Translog User is that the MT output has to first be pasted into the target text window The translator then commences post-editing (see Figure 4 6)



Saving Files and Autosave

To save your file by using the toolbar, select the Save button

You can also select the Save or Save As menu item from the File menu, or press the Ctrl S key

The editor prompts you for the name of the file to save

After you save your file once Epic can automatically save your file

The automatic save (*autosave*) function of the editor will save your file every 10 minutes

When you are creating a new file be sure to save your file right away

This causes Epic to start the autosave function.

The editor saves the files with an extension of ASV

The editor does not automatically overlay your original file

Einsparungdateien und speichern automatisch

um Ihre Datei indem Sie das toolbar zu sichern verwenden wahlen *Sie die* Speichertaste aus

Sie konnen die Abwehr oder außer als Menueinzelteil vom Datei- Menu auch auswahlen oder betatigen Sie die Ctrl S Taste

Der Herausgeber fordert Sie fur den Namen der Datei auf zu sichern

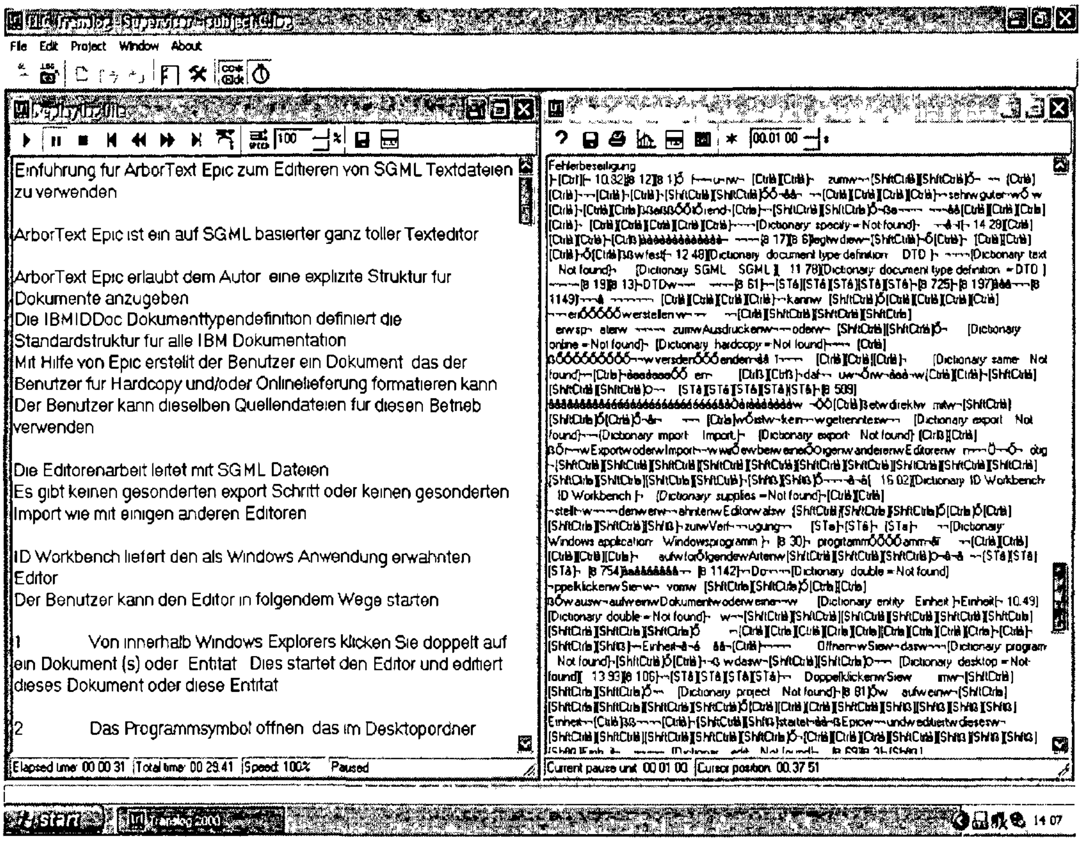
Nachdem Sie Ihre Datei einmal sichern kann Epic Ihre Datei automatisch sichern

Das automatische außer (speichern Sie automatisch Funktion des Herausgebers sichert Ihre Datei alle 10 Minuten

Wenn Sie eine neue Datei herstellen seien Sie sicher Ihre Datei sofort zu sichern

Figure 4 6 The Translog User Post-Editing Interface

The log file produced by Translog User can be viewed only using Translog Supervisor When the log file is opened, two parallel windows appear, one containing a “Linear Representation” of the target text production (on the right-hand side in Figure 4 7) and one where text production can be replayed like a video (on the left-hand side in Figure 4 7) The Replay window allows the user to speed up or slow down the replay, to jump forwards or backwards in the text, to pause and to save the final text product as a text file If the user clicks on the Linear Representation window at any time, the cursor will appear in exactly that part of the text that is currently being replayed in the Replay window In theory, this allows the analyst to map text production with mouse movements, keyboard activity, pauses and hesitations etc



Einfuhrung fur ArborText Epic zum Editieren von SGML Textdateien zu verwenden

ArborText Epic ist ein auf SGML basierter ganz toller Texteditor

ArborText Epic erlaubt dem Autor eine explizite Struktur fur Dokumente anzugeben
Die IBMIDDoc Dokumenttypendefinition definiert die Standardstruktur fur alle IBM Dokumentation
Mit Hilfe von Epic erstellt der Benutzer ein Dokument das der Benutzer fur Hardcopy und/oder Onlinelieferung formatieren kann
Der Benutzer kann dieselben Quellendateien fur diesen Betrieb verwenden

Die Editorenarbeit leitet mit SGML Dateien
Es gibt keinen gesonderten export Schritt oder keinen gesonderten Import wie mit einigen anderen Editoren

ID Workbench liefert den als Windows Anwendung erwahnten Editor
Der Benutzer kann den Editor in folgendem Wege starten

1 Von innerhalb Windows Explorers klicken Sie doppelt auf ein Dokument (s) oder Entitat Dies startet den Editor und editiert dieses Dokument oder diese Entitat

2 Das Programmsymbol offnen das im Desktopordner

Figure 4 7 The Replay and Linear Representation Windows in Translog Supervisor

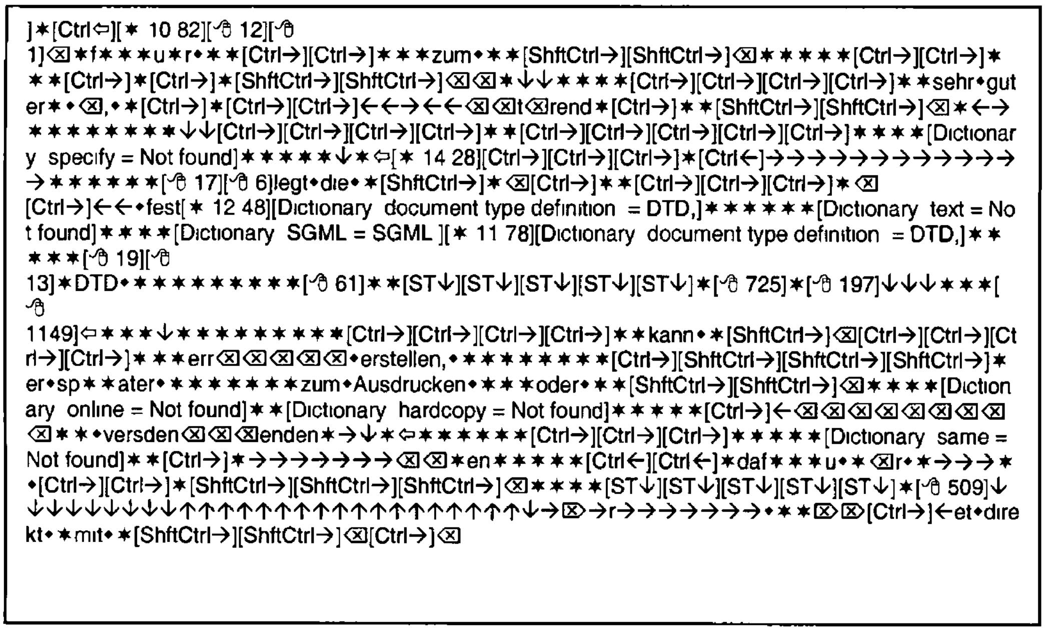
4.5.2 Strengths and Weaknesses of Translog

Translog is one of the few tools the author is aware of that allows translation researchers to record and analyse the translation (or post-editing) process as it happens on screen The software is easy to use and translators who are used to using a general word processor would feel unchallenged by Translog One of the main benefits of Translog is that it allows one to record and play-back, at different speeds, the translation text production process This gives an insight into *how* a text is produced and can be used to supplement data on the *end product*

Jakobsen (1999) acknowledges that the use of Translog could have an effect on the translation process itself However, his subjects reported that using Translog was very similar to writing an ordinary translation Lack of familiarity with the editing environment could also pose a problem, but this is easily overcome by familiarisation training in advance of experimentation It was the current author's experience also that training in advance meant that subjects were comfortable with the Translog environment

Unfortunately, there are certain weaknesses associated with the version of Translog used in this study (Version 1 0, Beta 4) The Translog program has been designed with translation in mind, which means that the target text window is usually blank to start with When post-editing, however, it is necessary to paste the MT output into the target window before post-editing can commence However, this cannot be done until the green "Go" button is pressed to commence logging The first couple of seconds are therefore taken up with pasting the MT output into and scrolling to the top of the target text window This is, admittedly, only a minor problem and simply means that the first few seconds of logging post-editing activity should be ignored

In addition, the symbols contained in the linear representation file are not displayed correctly under certain versions of the Windows Operating System (e g Windows XP) To rectify this problem, the user has to save the linear representation file as a "Rich Text Format" (RTF) file and display it in Microsoft Word in order to interpret it properly This detracts from Translog's feature whereby the user is normally able to observe the activity in the replay window and monitor the position in the linear representation file at the same time Table 4 1 contains the linear representation symbols used in Translog and a description for each one Figure 4 8 shows a sample of a linear representation file saved as RTF



```
]✱[Ctrl⇦][✱ 10 82][🖰 12][🖰
1]⌫✱f✱✱✱u✱r•✱✱[Ctrl→][Ctrl→]✱✱✱zum•✱✱[ShftCtrl→][ShftCtrl→]⌫✱✱✱✱✱[Ctrl→][Ctrl→]✱
✱✱[Ctrl→]✱[Ctrl→]✱[ShftCtrl→][ShftCtrl→]⌫⌫✱↓↓✱✱✱✱[Ctrl→][Ctrl→][Ctrl→][Ctrl→]✱✱sehr•gut
er✱•⌫,•✱[Ctrl→]✱[Ctrl→][Ctrl→]←←→←←⌫⌫t⌫rend✱[Ctrl→]✱✱[ShftCtrl→][ShftCtrl→]⌫✱←→
✱✱✱✱✱✱✱✱↓↓[Ctrl→][Ctrl→][Ctrl→][Ctrl→]✱✱[Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→]✱✱✱✱[Dictionar
y  specify = Not found]✱✱✱✱✱↓✱⇦[✱ 14 28][Ctrl→][Ctrl→][Ctrl→]✱[Ctrl←]→→→→→→→→→→→→
→✱✱✱✱✱✱[🖰 17][🖰 6]legt•die•✱[ShftCtrl→]✱⌫[Ctrl→]✱✱[Ctrl→][Ctrl→][Ctrl→]✱⌫
[Ctrl→]←←•fest[✱ 12 48][Dictionary  document type definition  = DTD,]✱✱✱✱✱✱[Dictionary  text = No
t found]✱✱✱✱[Dictionary  SGML = SGML ][✱ 11 78][Dictionary  document type definition  = DTD,]✱✱
✱✱✱[🖰 19][🖰
13]✱DTD•✱✱✱✱✱✱✱✱✱✱[🖰 61]✱✱[ST↓][ST↓][ST↓][ST↓][ST↓]✱[🖰 725]✱[🖰 197]↓↓↓✱✱✱[
🖰
1149]⇦✱✱✱↓✱✱✱✱✱✱✱✱✱✱[Ctrl→][Ctrl→][Ctrl→][Ctrl→]✱✱kann•✱[ShftCtrl→]⌫[Ctrl→][Ctrl→][Ct
rl→][Ctrl→]✱✱✱err⌫⌫⌫⌫⌫•erstellen,•✱✱✱✱✱✱✱✱[Ctrl→][ShftCtrl→][ShftCtrl→][ShftCtrl→]✱
er•sp✱✱ater•✱✱✱✱✱✱✱✱zum•Ausdrucken•✱✱✱oder•✱✱[ShftCtrl→][ShftCtrl→]⌫✱✱✱✱[Diction
ary  online = Not found]✱✱[Dictionary  hardcopy = Not found]✱✱✱✱✱[Ctrl→]←⌫⌫⌫⌫⌫⌫⌫⌫
⌫✱✱•versden⌫⌫⌫enden✱→↓✱⇦✱✱✱✱✱✱✱[Ctrl→][Ctrl→][Ctrl→]✱✱✱✱✱[Dictionary  same =
Not found]✱✱[Ctrl→]✱→→→→→→→⌫⌫✱en✱✱✱✱✱[Ctrl←][Ctrl←]✱daf✱✱✱u•✱⌫r•✱✱→→→✱
•[Ctrl→][Ctrl→]✱[ShftCtrl→][ShftCtrl→][ShftCtrl→]⌫✱✱✱✱[ST↓][ST↓][ST↓][ST↓][ST↓]✱[🖰 509]↓
↓↓↓↓↓↓↓↓↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↑↓→⌦→r→→→→→→→•✱✱⌦⌦[Ctrl→]←et•dire
kt•✱mit•✱[ShftCtrl→][ShftCtrl→]⌫[Ctrl→]⌫
```

Figure 4 8 The Translog Linear Representation Format

Symbol	Explanation
✱	Pause of 1 second or less
[✱ 10 82]	Extended pause with length of time
[Ctrl⇦]	Jump one word to left
[Ctrl→]	Jump one word to right
[🖰 12]	Mouse movement
[ShftCtrl→]	Select a word to the left
[ShftCtrl←]	Select a word to the right
⌫	Delete to left
↓	Move down one line of text
←	Move one character to the left
→	Move one character to the right
↑	Move up one line of text

Symbol	Explanation
angezeigt•werden	Typing of these exact characters
•	Insert blank space
[Dictionary specify = Not found]	Unsuccessful dictionary look-up for the word “specify”
[Dictionary document type definition = DTD]	Successful dictionary look-up for the term “document type definition”, equivalent given “DTD”
[ST↓]	Scroll source text down
⇦	Backspace

Table 4 1 Translog’s Linear Representation Symbols

The problems mentioned above are “bugs” which the developers of Translog are aware of and which, time-permitting, could be rectified At the time of this research, however, no fixes were planned and the software was used “as is”

4.6 Choice Network Analysis as a Research Methodology

Campbell (1999, 2000a, 2000b), Campbell and Hale (1999) and Hale and Campbell (2002) are interested in the concept of text difficulty in translation (or “translatability”) and, more specifically, whether or not a source text represents the same level of difficulty for translators who translate into different, unrelated target languages In Campbell (1999 38), it is acknowledged that one methodology for exploring source text difficulty is TAP However, the author points to the “well-known drawback” of TAP, i e that it is “not [ ] able to access unconscious processes” The alternative or complementary method proposed by Campbell and Hale is called “Choice Network Analysis”, or “CNA” for short In Campbell (2000b), CNA is explained in the following manner

> *Choice Network Analysis compares the renditions of a single string of translation by multiple translators in order to propose a network of choices that theoretically represents the cognitive model available to any translator for translating that string The technique is favoured over the think-aloud method, which is acknowledged as not being able to access automaticized processes*
>
> *(ibid 215)*

Thus CNA is presented as a method for constructing models of the mental processing underlying translation and it is also useful for estimating the relative difficulty of parts of source texts, where the measure of difficulty can be established based on the complexity of choices available to the translator Figure 4 9 provides an example of what a choice network diagram might look like for the translation into German of the heading “Introduction to using Arbortext Epic for editing SGML text files” [51]

[51] This choice network map was created using translations created by six students in a pilot post editing experiment

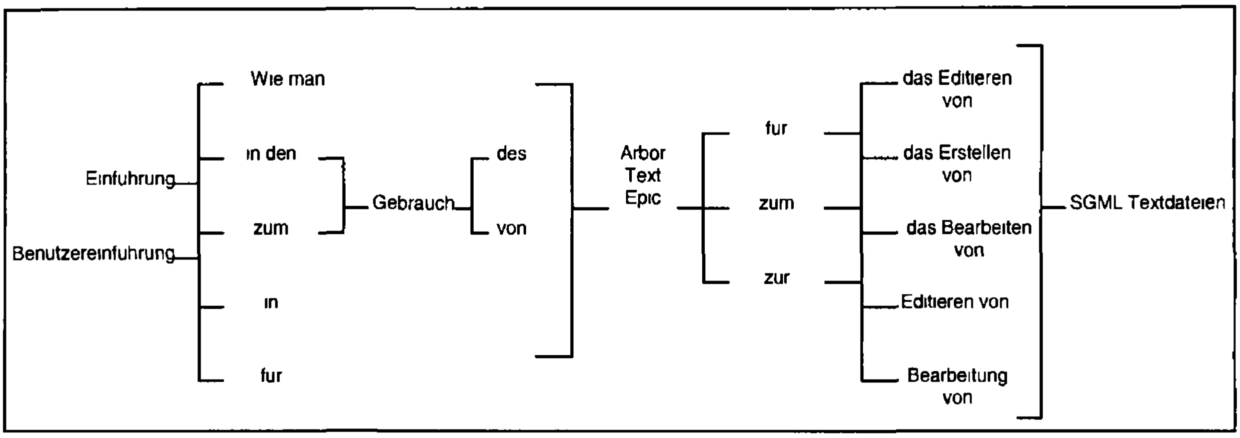


Figure 4 9 Example of a Choice Network

Campbell (1999) gives the following description of how CNA works

> *The main source of evidence about the amount of cognitive processing used is the mean number of alternate renditions made by a group of subjects translating the same item, where (on the evidence of the subjects' offerings) a single choice is offered as the translation of an item, it is assumed that the item requires minimal processing and is therefore easy to translate Where each subject offers a different rendition, then we assume that the range is available in principle to all the subjects and that large processing effort is required by each individual to make a choice from the range, the item is then a difficult one to translate*
>
> *(ibid 39)*

4.6.1 Application of CNA to date

Campbell and Hale (1999) use CNA to address two questions (1) Can the translation difficulty of English source texts be assessed?, (2) Does a given English text cause equal difficulty when translated into different languages? For this experiment, they draw on two sources for indirect assistance, i e the field of readability research and the field of text type research However, they point out that measures of readability, such as the FOG index, use crude criteria and have been seriously criticised The second source of potential assistance in answering the questions posed is the cataloguing of lexical and grammatical features of a large number of text types and the relationship these hold to readability However, they point out that "the complexity of an English text for monolingual reading may not equate to its complexity in reading for translation" (ibid 1) They suggest that text difficulty is only one of three variables to be considered in exploring translation difficulty "the others being translator competence and the translation task type" (ibid 2)

In Campbell and Hale's study, the source text was translated into three unrelated languages, Spanish, Arabic and Vietnamese Portions of text were judged to be "difficult" if subjects provided different renditions of a chunk of text If a chunk caused a similar level of difficulty in all three languages, it was judged to be "universally difficult" (although the authors concede that the use of the word "universal" may be a little ambitious)

Their analysis focused on lexis and grammar and isolated five areas of difficulty:

- Words low in propositional content (e.g. "become free from all opoid use")
- Complex noun phrases (e.g. "methadone treatment")
- Abstractness (e.g. "practice", "action")
- Official terms (e.g. "Anti-Discrimination Board")
- Passive verbs

Campbell and Hale found that English source text difficulties were common to the three target languages. They also found evidence for two "loci of difficulty", the first of which is comprehension where, they claim, difficulty is likely to be fairly universal. The second locus is production where, they claim, there would be different levels of difficulty depending on the lexis and grammar of the target language.

On the question of the practicality of CNA, they suggest source text "items" could be weighted according to their difficulty. "Harder" items would carry higher weights and different "item types" would have different weight assignments.[52] For example, on the evidence presented by Campbell and Hale (1999), the occurrence of passive verbs and complex noun phrases in a source text would be given high weights. The total text difficulty would then be based on the summing of the weights. This approach, although more simplistic, has obvious similarities with the approaches of controlled language scholars who have sought to measure text translatability in the past (see Gdaniec 1994, Bernth 1999a, 1999b, 2000, Underwood and Jongejan 2001; Bernth and Gdaniec 2001 and our discussion in Chapter 2).

Campbell (1999) outlines another experiment using CNA. The aims of this study were threefold - to establish (1) whether the source text is an independent source of difficulty in translation; (2) whether such difficulty is common to typologically different languages; and (3) possible reasons for text difficulty at the level of lexis (ibid: 38). Aims (1) and (2) are identical to the aims of the experiment outlined in Campbell and Hale (1999), whereas the third aim introduces a new dimension to the study. In this experiment, Campbell uses second language speakers of English who resided permanently in Australia and were undergoing professional translator training at the time of the study. They were asked to translate 250 words of text under pressure of time.

The case study proceeds on the basis that text difficulty is related to the processing effort needed to translate particular items in a text. It is important to note for the present study that the criterion of difficulty is not associated with the idea of correctness. Campbell

[52] The term "item" is not explained by Campbell and Hale in grammatical terms.

focuses on the decision-making process rather than on evaluating the product In addition to the mean number of alternative renditions, Campbell also assesses the extent to which subjects edit their work He assumes that an item that requires target-text editing is more difficult than an item that does not (ibid 39)

For the data analysis, investigators arbitrarily divided texts into analysable segments "generally on the basis of major constituent boundaries" (ibid 40) However, they then discovered that differences in translation occurred on a more micro level and so data were divided into smaller "chunks" Campbell acknowledges that this runs counter-intuitive to the notion that translators handle stretches of text of various sizes, but he hoped that larger stretches of difficulty would be revealed when the segments were reconstituted and this, he says, was largely borne out by the data

The number of alternatives for each chunk was then counted and converted to a raw score For example, if there were seven versions over nine subjects, then the score was 77 77 (7/9) Next, the raw scores were converted to z-scores (with a mean of zero and a standard deviation of one) The raw scores were then correlated language to language The scores for each chunk in each language were correlated with the number of edits Campbell's findings are summarised in the following statement

> *The results suggested that since common difficulties were encountered across subjects, texts could be said to be inherently difficult to translate, at least in the on-line mode Furthermore, it was found that considerable commonality of difficulty could be found across target languages [ ] It was found that less meaningful words were harder to translate, as were complex noun phrases and abstract nouns*
>
> *(ibid 57)*

In Campbell (2000b), CNA is used to investigate the phenomena of cross-clause ellipsis and relative clauses in translation from Arabic into English He chooses cross-clause ellipsis for his study specifically because it is "the maximum span across which relatively complex grammatical information must be retained in memory" (ibid 216) Although only four strings involving cross-clause ellipsis were found in the data, Campbell observes five different behaviours over nine subjects The behaviour is characterised according to whether subjects chose low risk, moderate risk, or high risk strategies to produce three possible results The following CNA diagram (ibid 220) illustrates how such a choice network can look

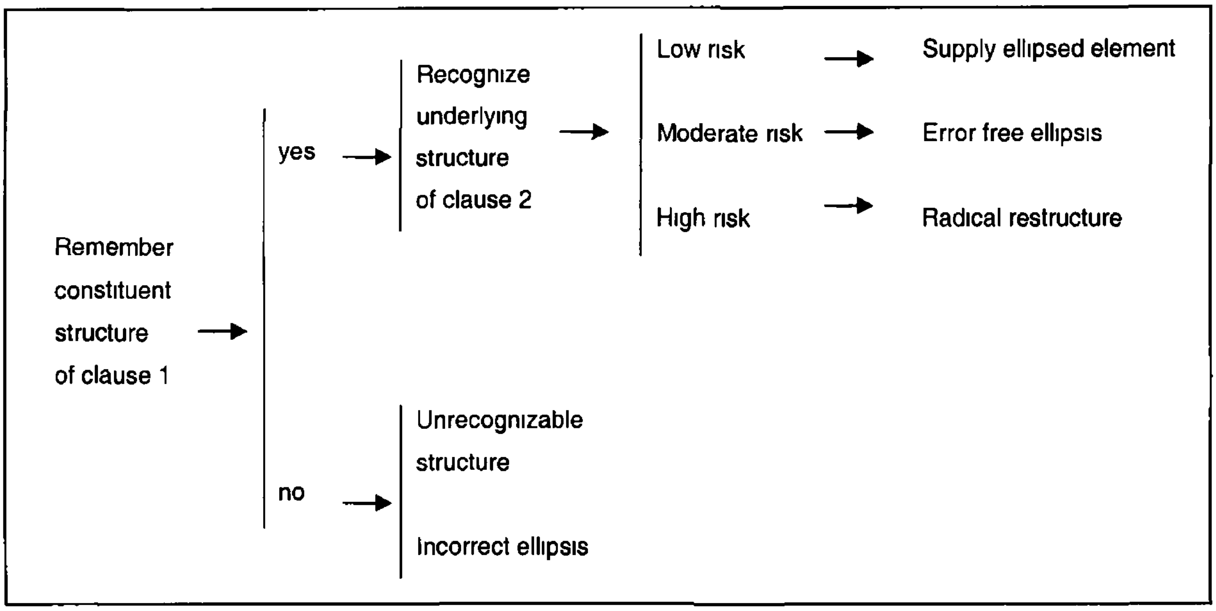


Figure 4 10 Choice Network for Translating Cross-Clause Ellipsis

Campbell concludes from this study that "information about cross-clause processing and about cognitive style or disposition can be gleaned from two or three critical structures" (ibid 227) He also suggests that this type of experimentation can equip translator educators with information about students so that the latter can be placed in appropriate instructional grades with appropriate syllabi and objectives

In Hale and Campbell (2002 16) it is acknowledged that previous studies by them used "a rather broad brush to paint a picture of difficulty" Consequently, in their 2002 paper they set out to refine the approach and to include a discussion on the relationship between difficulty and accuracy In their analysis they concentrate on the ST items official terms, complex noun phrases, passive verbs and metaphors, all of which were found to produce a high number of different renditions in their previous research They include 20 subjects in their experiment, 11 of which were Spanish speakers and nine of which were speakers of Arabic All subjects were asked to translate two texts, of approximately 260 words each, from English into their mother tongues The choices made by the subjects were then analysed by Hale and Campbell (ibid) While they found that all item types under analysis "caused quantitatively similar difficulty to the subjects" (ibid 28), the causes of difficulty were qualitatively different So, for example, when faced with translating an official term, subjects grappled with the competing strategies of explicitation, transcription and nativisation whereas when faced with the translation of a metaphor, subjects had trouble choosing competing lexical items Hale and Campbell conclude that the correlation between difficulty and accuracy is highly complex, noting also that the notion of "accuracy" is difficult to define They then propose that accuracy can be achieved "by striving for pragmatic equivalence, where the criterion of appropriateness is a major indicator" (ibid 29)

4.6.2 Strengths and Weaknesses of Choice Network Analysis

The studies of Campbell and Hale using CNA have led to interesting observations and conclusions regarding text difficulty for translation Specific linguistic *phenomena*, e g passive verbs, official terms, complex noun phrases etc , have been identified as presenting difficulty in translation from English into unrelated target languages Some practical uses of this methodology have also been suggested by Campbell and Hale, e g the classification of translation students into groups according to capability Additionally, CNA does not entail any of the drawbacks associated with TAP as outlined in the section above

Choice Network Analysis does, however, have some weaknesses Firstly, CNA appears to be a relatively new and untested methodology (tested only, it would seem, by Campbell and Hale themselves) Secondly, Campbell and Hale (1999) mention that text difficulty is only one of the variables to be considered in the measurement of translation difficulty The other variables, translator competence and the translation task, are mentioned by them but not elaborated upon Nor is the question addressed of how an individual translator's style affects choice in translation Thirdly, although Campbell (2000a) mentions that choices made at any point in a string may constrain subsequent choices due to the grammar of the target language, no recommendation is made on how to account for this in the analysis of a choice network diagram Finally, the maximum number of subjects used in the experiments described above was nine The issue of group size is not addressed by Campbell and Hale This leads to an inevitable question how many subjects are required to produce all possible renditions of a ST chunk? If the group size is small, then, in theory, other renditions are possible As many researchers of translation will testify, it is possible for one hundred translators to produce one hundred different renditions of the same source sentence Thus it seems that data regarding the cognitive load for a particular chunk, based on CNA alone, can never be deemed to be "conclusive"

Despite these weaknesses, the inclusion of CNA as a research method in the current study is desirable As Campbell (2000a) points out, CNA can be used to generate hypotheses that can be tested by other research methods CNA generates models that can be made as simple or as complex as the hypothesis under investigation demands Since TAP has been ruled out as a method for assessing the cognitive load in post-editing and its correlation with the difficulty or translatability of the source text, it would seem that CNA should be used as a method to complement the use of text monitoring software, as discussed above This is in line with the recommendations made by translation process researchers that triangulation is important in TPR The inclusion of CNA as a research methodology also overcomes the criticism levelled by Toury (1991 47) against translation process research that the final product is often completely neglected Finally, it is hoped that the use of CNA throughout this study will put the current author in the position of being able to comment on this method and its usefulness for translation research and that this commentary will be a valuable contribution to the information available so far on Choice Network Analysis

4.7 CONCLUSION

In this Chapter we introduced the main topics in Translation Process Research Their relevance for research into post-editing effort was considered and it was found that some topics raise important questions for the current study Those questions will be dealt with in more detail in Chapter 5

The core data source used in TPR, i e Verbal Report Data as recorded in TAPs, was also discussed here The controversial nature of TAP was highlighted and we considered whether or not TAP was an appropriate methodology for the current study, concluding that the drawbacks mentioned by many researchers would have a negative impact on the research objective Alternative methodologies, i e text production monitoring software and Choice Network Analysis, were then described Although there are weaknesses in both methodologies, it is clear that, if used in triangulation, they have the potential to provide us with the necessary tools by which we can investigate the correlations between translatability and post-editing effort A useful methodological framework thus begins to emerge Chapter 5 will discuss this methodological framework in more detail

Chapter 5

5. METHODOLOGY

5.1 INTRODUCTION

In the preceding chapters, we set the scene for the investigation of correlations, if any, between NTIs and post-editing effort Chapter 5 turns attention to the theoretical and practical methodological questions that arise The principles of research design expounded by Frey et al (1991) are drawn on in the first section where the topics of independent and dependent variables, operationalisation, units of analysis and internal, external and measurement validity are examined In the second section, we discuss the practical decisions made regarding research design and data analysis

In Chapter 4, we introduced Translog and explained our reasons for using it in this study Because the research involved the use of Translog for the first time by the researcher, it was decided that a pilot run of the experiment would be necessary A pilot run also helped to ensure that the risk of flaws in the research design was reduced to a minimum This was particularly important as we only had access to the professional translators who participated as subjects for a very limited time

For the pilot study we recruited students of translation who were native speakers of German and who were participating in one year's study abroad at Dublin City University The profiles of students who expressed an interest in participating were examined before they were recruited to the study (e g mother tongue, years spent studying translation, experience with word processing and translation tools) The pilot project enabled the researcher to make sure that all instructions were clear for participants and that any risks regarding the use of technology could be reduced, if not eliminated The pilot study is discussed in more detail below under Research Design

The Localisation Division of IBM in Germany (Stuttgart) agreed to allow some of their translators to participate in the main study Prior to selection, potential subjects were asked to fill in a short survey, which is discussed in more detail under Internal Validity The selection of subjects is also discussed here and in further detail under Research Design

Following the selection of subjects, the source document was prepared and sent to IBM for terminology extraction Krings (2001) found that the use of reference works is highly translator-specific He also found that reference work usage is more common with translation than with post-editing For these reasons the following measures were implemented

- Terminology identified by the MT system as not being in the dictionary was subsequently coded The identification of terms not included in the MT dictionary was carried out by IBM Germany while the coding of the dictionary was carried out by the researcher. 79 terms were coded altogether. When coding was complete, the dictionary file was returned to IBM Germany for use during machine translation. Prior to the study, subjects were informed that terminology had been coded so that they could have confidence in the proposed terms;

- The same terminology was then entered into the Translog dictionary so that, when in doubt, subjects could look up terms to make sure that they were correct. Translog recorded this action as a dictionary look-up in the log file for each subject;[53]

Once the MT dictionary had been coded, the source document was then machine translated into German and the preparation for the full study was complete.

The researcher travelled to IBM in Germany to conduct the study. The translators were requested to arrive at the allocated room at specific times on specific days. At that point, they were briefed in more detail about the study and were allowed to practise for some time with Translog until they felt comfortable with the working environment. The theoretical and practical questions underlying the research methodology are considered now in more detail below.

5.2 THE THEORETICAL VIEWPOINT

5.2.1 Independent and Dependent Variables

The current study wishes to examine the correlation between post-editing effort and the presence of specific negative translatability indicators (NTIs) in sentences. The assumption is that post-editing effort will be lower when there are no NTIs in a sentence and that it will be greater when NTIs are present. Therefore, the independent variable (ibid: 46) in this study is the negative translatability indicator, while the dependent variable (ibid) is post-editing effort.

5.2.2 Operationalisation

All research projects require an operational definition. This is defined by Frey et al. (ibid: 95) as a definition which describes the observable characteristics of a concept. An operationalisation must possess three characteristics. Firstly, it must be adequate, that is, it

[53] It emerged that terminology management in Translog was somewhat problematic. If a term was selected from a sentence and placed in the Translog dictionary, then that term was available to subjects if they decided to look for it in the dictionary at that specific point in the text. However, if the term occurred in another sentence and subjects tried to find it in the dictionary at that point in the text, then the dictionary would report that the term was unavailable. This perhaps had an influence on subjects' dictionary look-up behaviour, which is discussed in the chapter on Data Analysis.

must provide a complete description of the characteristics being observed It must also be accurate According to Frey et al (ibid) this means that the operational definition is universally agreed Thirdly, it must be clear for future readers and researchers Operationalisations can be both qualitative and quantitative A combination of both can be used to achieve triangulation

To establish an operational definition for the study at hand, we must first describe the observable characteristics of *translatability* and *post-editing* A description of observable characteristics can only be derived from the characteristics outlined by research published to date Thus, the observable characteristics of CL, in general, and translatability, specifically, have been extracted from the literature published to date (Gdaniec 1994, Kohl 1999, Bernth 1999a, Bernth and McCord 2000, Underwood and Jongejan 2001, Bernth and Gdaniec 2001) and these will be used as part of the operational definition In particular, we have drawn on the translatability rules expounded by Bernth and Gdaniec (2001) in order to create a list of NTIs The observable characteristics of post-editing have also been extracted from literature published to date (Krings 2001, Senez 1998a, 1998b, Loffler-Laurian 1981, 1984, Vasconcellos 1986 etc) and these have been used as part of the operational definition We have seen that researchers differentiate between different types of post-editing, e g partial and full In this study we are interested in full post-editing and there are a number of reasons for this one of Krings's suggested measures of post-editing effort (in addition to temporal, technical and cognitive effort) is Relative Post-Editing Effort, which involves a comparison of full post-editing effort with human translation This presented one reason for focusing on full post-editing effort Also, limitations on the scope of this research project would not permit a full investigation into both full and partial post-editing effort This is, of course, a topic that could be researched in the future

5.2.3 Unit of Analysis

"Attention units" are identified by Jaaskelainen (1987, 1990) as appropriate units of analysis in translation process research These units are marked by a shift in the translator's focus of attention as reported in the translator's think-aloud protocol Since TAP will not be the method of choice for this study, it will not be possible to measure the unit of analysis by a verbally reported shift in focus Seguinot (1989b) and Krings (2001) both identify units using pause and hesitation phenomena For Krings a pause is a break in processing of one second or more Since Translog is capable of measuring pauses in seconds, pause and hesitation phenomena could be used to identify units of analysis in this study However, given that pauses are reported to be individual (Hansen 2002) and erratic in nature (Alves 2006), we could then be faced with analysing units of different length across post-editors Given that we are interested in measuring the effect on post-editing of the presence of NTIs in a sentence, this argues in favour of using the sentence as the unit of analysis as it is relatively easy to construct sentences that either contain or do not contain specific NTIs "Sentence" is generally understood here to mean a string of words, usually starting with a capital letter and ending with a punctuation mark Although they do not conform to this definition, an item in a bulleted list or a heading would also be treated as an independent unit by a CL checker Therefore, the most appropriate term to use is "segment", which covers the concept of "sentence", as defined above, but also covers a heading or bulleted list item Therefore, the units of analysis in this study will be source text segments and they will be classified either as having NTIs (termed "$S_{(nti)}$ segments") or has having "minimal" NTIs (termed "$S_{(min\ nti)}$ segments") Given the fact that many words can be polysemous or have multiple parts-of-speech (POS) and that these features can also be counted as NTIs, it is impossible to state that any one segment has "no" NTIs Hence the word "minimal" is used to denote a segment where all NTIs, with the exception of polysemy and multiple POS, have been removed While polysemy and multiple POS are not included in our list of NTIs to be tested, any problems arising from either of these NTIs during machine translation or post-editing will, of course, be noted

5.2.4 Validity

Research projects can demonstrate validity according to a variety of criteria Frey et al identify the different types of validity as internal, external, and measurement

5.2.4.1 Internal Validity

Internal validity refers to the accuracy of the conclusions drawn Threats to internal validity include threats due to researchers, threats due to how research is conducted, and threats due to research subjects (ibid 125)

Threats Due to Researchers

Two possible threats to internal validity can be posed by researchers The first is called the "researcher personal attribute effect" Frey et al report on research which demonstrates that different researcher characteristics (e g race, gender) influence subjects' responses This is likely to occur under two conditions when the research task is ambiguous such that subjects look to the researcher for information on how to perform, and when the research task is related to the personal characteristics of the researcher (e g colour, creed)

The second effect is known as the "researcher unintentional expectancy effect" This occurs when researchers influence subjects' responses by unwittingly revealing the type of results they desire

To control for both of these effects, Frey et al recommend that the researcher can remove him or herself from the study and employ a wide variety of research assistants who are blind to the objectives of the research An alternative suggestion is to follow standard procedures so that everyone is exposed to the same research environment

Employment of a wide variety of research assistants for the study at hand was not feasible Therefore, the second method for neutralising the researcher effects mentioned above have been employed, i e the research environment was kept constant and a script was prepared for use by the research facilitator to give identical information and instructions to subjects This script was written in such a way so that the desired or expected research results were not communicated, either intentionally or unintentionally, to the subjects See Appendix B for a copy of the script and the instructions given to the post-editors and translators

THREATS DUE TO HOW RESEARCH IS CONDUCTED

The second threat to internal validity identified by Frey et al is that which is posed by how the research is conducted The factors contributing to this effect include the validity and reliability of the procedures used, history, sensitisation and data analysis The first of these factors requires accurate and consistent application of the research procedure This can be achieved by using accurate measurement techniques in a consistent manner and by ensuring that all subjects are exposed to the same levels of variables in a consistent manner To minimise these threats for the current study, all subjects were given the same assignment, with the same text and they were allocated the same amount of time to complete the assignment Completion of the assignment was carried out using the same technology and workstation, which was located in the same room throughout the data collation cycle

The second factor, history, refers to all changes in the environment external to the study which may influence a subject's behaviour This is particularly important for longitudinal research Since the current study was designed to collect data at one period in time only (i e over three days in one week), this factor was ignored

When analysing data, the analyst may be influenced by an initial measurement or procedure when carrying out a subsequent measure or procedure and this is known as "sensitisation" Krings, for example, reports that the evaluators in his study became more "lenient" towards poor MT output as they progressed through the texts (Krings 2001 263) For the study at hand, two of the measurements, i e technical post-editing and temporal post-editing effort, are quantitative in nature and involved counting the number of words changed, deleted, inserted etc as well as the amount of time required to implement these edits Given the objective and numerical nature of this data analysis, sensitisation was not a threat For the third measurement - Choice Network Analysis - the post-editing products were compared with a view to identifying specific parts of segments that had been changed and, again, sensitisation was not considered to be a threat

The last risk factor for how research is conducted is identified as data analysis When improper procedures are used to analyse data, internal validity is threatened To minimise the threats from this risk, data were analysed in a systematic way and results were recorded immediately in spreadsheets The data analysis procedures are described in more detail in the section on Research Design

THREATS DUE TO RESEARCH SUBJECTS

Frey et al identify a number of threats to internal validity due to research subjects the Hawthorne effect, selection, statistical regression, mortality, maturation, and intersubject bias

The Hawthorne effect occurs when people change (i e improve) their normal behaviour because they are aware that they are being studied One way to control for this is not to let people know that they are being studied However, this raises ethical issues and is not a suitable solution for the current study As will be discussed in our chapter on Data Analysis, we witnessed changed behaviour during our experiment, but this change in behaviour was a disimprovement in performance rather than an improvement one subject changed his/her post-editing behaviour after segment 50 while another changed his/her behaviour at segment 130 Both subjects appear to have decided that they would not fix any more errors in the text and produced a target text that was incomprehensible and unusable The result was that we had to exclude the data from the first subject completely as they were invalid For the second subject, the behaviour change occurred later in the text and this contributed to a decision not to proceed with the analysis beyond a certain point This will be discussed in more detail in the chapter on Data Analysis For the moment we will draw attention to the fact that the threat of a change in behaviour is a very real one Measures for alleviating this threat for studies similar to this one are suggested in Chapter 6 – Data

Analysis However, given that the behaviour reported above was unexpected, those measures were not implemented for this study [54]

The subjects and texts selected for any study bring with them a variety of different individual characteristics and skills and each one may affect the research results Researchers have commented on the difficulty of recruiting suitable subjects for translation process research in particular (Krings 2001, Jaaskelainen 1987, 1990) As mentioned in Chapter 4, when studying the differences between professional translators and non-professional translators, Jaaskelainen used what Krings termed "semi-professional translators" (i e 5th year third level students) as professional translators The students' lack of exposure to a professional working environment would most likely have influenced Jaaskelainen's results The same challenges were faced by this study firstly, a decision had to be made as to whether professional translators or semi-professional student translators were recruited, secondly, a decision was required as to whether or not the translators should be experienced in post-editing The most desirable context would have been to have a pool of professional translators, some with and some without post-editing experience as this would be most representative of the current professional situation However, the possibility of finding such a group of potential subjects was limited not only by individuals' availability, but also by budget, and by the fact that not many translators have post-editing experience The final make-up of the group depended on the availability of subjects The most important point was to ensure that the group was as homogenous, in terms of skills and experience, as possible IBM Germany kindly allowed some of their translators to participate in the study The number of translators available was 12

The 12 translators were asked to fill out a questionnaire prior to being accepted as research subjects (See Appendix C) The questionnaire had multiple objectives, i e

- To make sure that all subjects were native speakers of German
- To make sure that all subjects had undergone translator training
- To make sure that all subjects had spent three or more years working as a professional translator
- To make sure that all subjects regularly translated from English as a source language
- To make sure that all subjects regularly worked in the IT industry and were, therefore, familiar with the domain, terminology and text type

[54] Note that all subjects were being paid for their time by their employer and all gave the impression that they were happy to participate in the study and interested in the outcome For this reason, the modification in behaviour by P5 and P6 came as a surprise

- To make sure that all translators had experience with word processors and translation tools
- To ascertain subjects' previous experience with MT, post-editing, and CL
- To ascertain subjects' opinions regarding MT If any subjects had demonstrated strong negative opinions about MT they would have been excluded as it was felt that this might bias their post-editing performance

Once all questionnaires had been returned, the researcher was satisfied that the subjects constituted an adequately homogenous group in terms of qualifications, professional experience and experience with computers, the text type and translation tools Of the 12 subjects, only one reported having directly used an MT system prior to the study (and this was on only one occasion) Six subjects reported previous experience with post-editing, ranging from a 20,000 word test to a 100,000 word project Six reported having no experience with post-editing When asked to select a statement out of the four statements listed below that most accurately reflected their opinions of MT, all 12 subjects selected statement 2, i e *I think that MT can be used in restricted and controlled circumstances*

I think MT is useless and I would never use it

I think MT can be used in restricted and controlled circumstances

I think MT is quite good and would consider using it

I think MT is of great benefit to translators and translation clients

Once subjects had been checked according to the criteria above, it was felt that the threats due to subject selection had been minimised as much as possible

Mortality refers not only to the loss of subjects from the beginning to the end of a research cycle, but also to the loss or destruction of documents or data The risk of data mortality was high since technology played a large part in the experiment and is notoriously unreliable To control for this risk, instructions were included in the script mentioned above for the regular backing-up of data Data were saved to a number of locations immediately after the conclusion of each subject's post-editing session In addition, a strict naming convention was employed for files involved in each experiment to reduce the risk of file mix-up

Maturation refers to internal changes that occur within people over the course of time Due to the limited time-frame over which this experiment was conducted, maturation posed no risk

Finally, intersubject bias occurs when those being studied influence each other There was certainly a risk with the current study that subjects might talk to and influence each other if they were given the opportunity to do so To control for this, subjects were briefed individually and the experiments were conducted on an individual basis Subjects were given specific time-slots in which to arrive at the room Most subjects were not working together in the same building and, therefore, had little opportunity to discuss the study with one another In addition, the briefing document specifically asked them not to discuss the study with other participants The staggering of time slots also helped towards controlling the situation with regard to potential technological problems, which could have been more problematic if a number of people were working in the room at the same time

5.2.4.2 External Validity

External validity refers to the generalisability of the findings Three factors influence the extent of external validity sampling, ecological validity and the replication requirement

Sampling

Sampling is related to the issue of subject selection, discussed above Of the methods of sampling proposed by Frey et al , the two methods most relevant to this study are "purposive sampling" and "network sampling" Purposive sampling involves selecting subjects non-randomly because they possess certain characteristics or skills Network sampling is also known as the "snowball technique" This applies to a situation where subjects are asked to refer researchers to other people who could serve as subjects As discussed above, a purposive sampling approach was adopted here in order to recruit subjects with the right qualifications and experience

Ecological Validity

Ecological validity centres around the need to conduct research so that it reflects real-life situations Frey et al put it like this

> *Studying communication behaviour in natural settings increases the generalizability of research because communication processes may be thought of as streams of behavior, and like all streams, their course is shaped by the terrain through which they flow If we ignore the banks and study only the stream or divert the stream into an artificial container and study it there, our knowledge of that stream is inevitably limited*
>
> *(ibid 136)*

The ideal scenario for the current study might have been to use experienced post-editors in their normal work environment This, however, was not possible, because the normal work environment includes many distractions, whether that be from people coming to one's desk or from telephone calls The data collection stage of this study required full concentration on a text processing task for a certain period of time and this, in turn, necessitated the removal of subjects from their normal work environment and their placement in a controlled laboratory environment Although translators were not in their usual work environment, they were still located in an office, in a building that was familiar to them Frey et al point out that research can still be ecologically valid even if it does not take place in real-life circumstances, if attention is paid to procedures for enhancing ecological validity For this study, the following measures were implemented

- Each subject worked in an environment that was similar to any modern translation department, i e open plan office space, with a number of computer workstations
- The Translog software runs in a normal Windows environment, which is the usual desktop environment for most translators
- The software uses the same text processing controls of the normal Windows environment (e g CTRL + C for cut, CTRL + V for paste etc), and the subjects were familiar with this
- A short familiarisation run with the software was carried out so that each subject felt comfortable with the working environment
- The subjects were familiar with the text type used and with the domain

One final issue regarding ecological validity must be addressed Kohn (1988) and Krings (1986) characterise the main stages of translation as being pre-processing, main processing and post-processing The current research concentrates on the events in the main processing stages of post-editing Since the focus here is on how NTIs affect post-editing effort, we have not included the monitoring of reference work usage, time required to read and research the text in advance, or proofreading Subjects were asked to concentrate solely on text production In addition, since full and, especially, partial post-editing usually occur in a time-restricted scenario, it is reasonable to assume that the post-editing process demands that less time be spent on terminology, parallel text and target audience research and that more time be spent on rapid text production In order to familiarise themselves with the process of post-editing and the Translog environment, subjects were allowed to do a practice run on a short, similar text in advance of the experiment

One final point to make on the subject of ecological validity is that of the twelve subjects, three were randomly selected to act as a control group These three subjects, referred to as "T1", "T2" and "T3", were asked to translate the text instead of post-editing it While the focus of the research was not to create a comparison between translation and post-editing processes and products and, while this meant reducing the amount of data on post-editing, this decision enabled a calculation of one of the important indicators of post-editing effort, i e *relative post-editing effort* (Krings 2001) It also meant that we could refer to the translation products when analysing the post-editing products during the Choice Network Analysis stage and this proved to be helpful in offering explanations for certain decisions made during post-editing (for more discussion of this, see the chapter on Data Analysis)

Replication

Only when the results of a study have been reproduced in several replications can the original study be confirmed and extended (a contention by Tukey 1969, reproduced in Frey et al (ibid 138)) To ensure the possibility of replication and, consequently, to further guarantee external validity, the information on the text used, the data analysis procedure, the experiment set-up, subject profile etc have been documented in detail

5.2.4.3 Measurement Validity

Measurement validity refers to the ability of a measurement technique to tap the actual meaning of concepts being investigated (ibid 199), while measurement reliability involves measuring a variable in a reliable and consistent manner (ibid 120) The former is assessed at a conceptual rather than a numerical level For example, for the study at hand, we have drawn on research to date in both CL and post-editing to ensure that we are measuring features of translatability and post-editing that are accepted as valid measurements of those two concepts

Measurement reliability, on the other hand, addresses how the occurrence of random errors (when participants make a mistake) and measurement errors (when researchers make a mistake) can be reduced The methods for reducing the occurrences of such errors proposed by Frey et al are pilot testing, questionnaires, interviews and observations In other words

> *Measurement validity and reliability can be increased by combining quantitative and qualitative measuring procedures in the same research study, a practice referred to as triangulation*
>
> *(ibid 124)*

A pilot test was implemented to check the validity of using Translog as a research tool (this will be discussed in more detail later) In addition, triangulation of methodologies (Translog and CNA) was used Finally, the use of data collection templates meant that the data were captured in a consistent manner for each segment (see Appendix D for an example)

5.3 Research Design

The previous section addresses the methodological framework from a theoretical point of view, taking operationalisation, measurement techniques and validity into consideration The theoretical aspects now need to be implemented in a practical research design

5.3.1 Relative and Absolute Post-editing Effort

As already mentioned, Krings (2001) differentiates between relative and absolute post-editing effort As a control measure for this study, human translation of the experimental text was recorded and compared to post-editing effort Therefore, it was possible to measure relative post-editing effort

5.3.2 Use of a Source Text

Krings maintains that the cognitive load is greater when the post-editor has to work with three texts instead of two However, the norm for post-editing is currently one where source, MT and target text are visible to the post-editor To maintain ecological validity, the source text was made available to the research subjects in Translog

5.3.3 Text Type

The text type chosen for this research came from the domain of IT and belonged to the genre "user manuals" The justification for this was

- the current demand for translation of these particular text types is relatively high,[55] also translation demand in this domain has been growing over the past decade and there is an interest in the possibility of using MT for the translation of this text type,
- the fact that little work has been published on CLs relating to this domain,
- the availability of such text types

The text chosen was selected from a user manual describing the use of the IBM internal authoring tool known as the ID Workbench The use of an IBM document was appropriate for a number of reasons

- the IBM MT system was to be used and one could reasonably expect that this system would be more tuned to IBM terminology than other MT systems,
- the translators would be familiar with the writing style,

[55] According to Arle Lommel survey analyst with the Localization Industry Standards Association (LISA), the localisation industry has experienced considerable growth over the last decade (with some difficulty in the late 1990s due to the heavy reliance on the IT industry) (personal communication, Arle Lommel 2006) A substantial majority of localisation companies involved in that growth translate user manuals 65% out of 255 respondents to LISA's 2006 survey report translating user documentation (Lommel and Ray 2006)

- IBM gave permission to use this text and their CL checking tool

The document selected for the experiment is both a Getting Started and User's Guide (Release 3 6, October 30, 2001) It consists of a Table of Contents, twenty-three chapters, appendices, and an index, all contained in 517 pages IBM's own controlled language rules were not applied to the text during production Given the scale of the current study, it was necessary to select a smaller section of text for the empirical investigation Upon consideration, a portion of Chapter 9 "Editing SGML Files with Epic" was selected This portion of text was selected because it displays many standard characteristics found in IT User Manuals, e g

- Descriptive passages

e g "ArborText Epic is an SGML-based text editor that allows you to specify an explicit structure for your documents "

- Instructive passages

e g "From within Windows Explorer, double-click on an IBMIDDoc document or entity "

- Numbered and bulleted lists

e g Drag and drop editing for easier moving and copying of text and tags

- Abbreviations

e g IBMIDDoc, SGML, IBM, OS/2, WYSIWYG, XHTML

- Hyphenated words

e g "SGML-based text editor"

- Noun Phrases

*e g "The **Insert Markup** dialog remains available for you to select additional tags "*

- Proper Nouns

e g Epic, Xyvision

- Menu names and items

*e g "To leave the editor, from the **File** menu, select the **Exit** item "*

- Program messages

e g "Your document is OK when you see the message `No completeness errors found` "

In addition, Chapter 9 also includes graphics (mainly screen-shots), cross-references, various heading levels, and basic formatting (bold, italic, font changes) [56]

5.3.4 Number of Words

When deciding on the number of words to be included, the criteria that needed to be considered were

- Coverage The text must demonstrate a suitable variety and number of linguistic problems in order to include a number of occurrences of each NTI, which, in turn, will allow for generalisations to be made about the relationship between post-editing and NTIs
- Time required/concentration capacity Account must be taken of the average length of time a translator might spend post-editing one text without taking a break It was assumed that two hours of uninterrupted text processing was likely to be the maximum one could demand of a translator without having to take lapses of concentration into account If we assume that a translator is expected to translate 2,000 words per day on average (researcher's own experience of the localisation industry), this would mean that the translator could translate approximately 250 words in one hour (assuming all terminology research etc was completed) Normally, one assumes that post-editing (both for information and gisting purposes) would occur at a faster speed (and we have seen claims to this effect in Chapter 3) If this is taken into account, we can make a conservative estimate that approximately 500 words could be post-edited in one hour, when the assignment is for full post-editing, and approximately 1,000 words could be post-edited in two hours The challenge this raised was to find a text of that length with an adequate occurrence of the two segment types ($S_{(min\ nti)}$ and $S_{(nti)}$) to enable generalisation of the results The final excerpt chosen had a total of 1,777 words [57]

5.3.5 Language Pair

The choice of source language was predetermined by the fact that most CL rules have been designed for English as a source language The literature on negative translatability indicators, in particular, focuses on English as the SL The decision regarding the target language depended on a number of factors (1) the researcher's ability to analyse it, which limits the choice to French or German, (2) the availability of a machine translation system for

[56] Price (1984) includes all of the features mentioned above in his guide on how to write a computer manual Also the master document contains all of the features mentioned by Byrne (2004 31 45) in his discussion of the features of a "user manual", i e a table of contents, titles, headings, overviews, tables, graphics and drawings lists abbreviations and acronyms and specialised terminology

[57] This is longer than 1 000 words because we wanted to select text down to a natural section break in the document

this language pair, and (3) the availability of post-editors All of these factors led to the selection of English-German as the language pair

5.3.6 Creating the Test Document for Post-Editing

The justification for the selection of the text type was given earlier Using the literature on translatability measurement (Gdaniec 1994, Bernth 1999a, 1999b, Bernth & McCord 2000, Underwood and Jongejan 2001), in particular, Bernth and Gdaniec (2001), where a list of translatability rules are offered, a list of NTIs to be investigated was created (see Appendix E) All of Bernth and Gdaniec's translatability rules, bar numbers 24, 25 and 26, were used as input to the list of NTIs to be investigated It was not possible to test the latter rules, due to their pragmatic, rather than linguistic, nature Some additional NTIs were added to the list, based on the error messages generated by the two CL checkers used, e g use of multiple prepositions or stand-alone pronouns with indefinite reference Altogether, this generated a list of 29 NTIs to be investigated

In order to test the relationship between specific NTIs and post-editing effort, it was necessary that the test document display a sufficient number and range of NTIs In its raw state, Chapter 9 did not meet this objective and so editing was necessary The method used for editing is described here

First, all segments were numbered in the test document The document was then submitted for analysis to two Controlled Language checkers, EEA and Sunproof, and the resulting error messages were recorded for each segment[58] The error messages can be viewed in Appendix F The NTIs identified by the two CL checkers were recorded on a segment-by-segment basis The number and type of NTIs found throughout the file were examined By examining these data, it was possible to see that some NTIs were over-represented in the file (e g the use of pronouns), while others did not occur at all (e g slang) The file was then edited to balance the frequency of different types of NTIs Further comments on the problems associated with this task are presented below Microsoft Word's "Track Changes" feature was used to record the changes made and the "Comment" feature was used to insert comments on the changes made Appendix G shows the test document before and after editing A comparison of the performance of the two tools would make for an interesting discussion, but this is beyond the scope of the current study

One of the challenges encountered while preparing the test document was to ensure that there was an adequate number of occurrences of each NTI in order to provide adequate data on the post-editing effort And yet this had to be achieved while taking the document

[58] Two tools were used in order to maximise the number of NTIs identified The researcher gratefully acknowledges Sun Microsystems for granting permission to use their CL checker, Sunproof

length, concentration levels of subjects, and time-frame of the overall study into account An additional, important consideration was to ensure that the text would still be recognised by the post-editors as a user manual as opposed to a list of artificially constructed sentences Every effort was made to ensure that each NTI identified by Bernth and Gdaniec was represented at least twice in the test document For some NTIs, this is the maximum number of occurrences, e g usage of slang It was necessary to reduce the frequency of occurrence in the document for other NTIs because they occurred very frequently, e g use of pronouns In addition, it was also necessary to have an adequate representation of $S_{(min\ nti)}$ segments (i e those with *minimal* negative translatability indicators) so as to compare the post-editing effort for the latter with segments containing known NTIs All of this had to be achieved within the small number of words contained in the test document (1,777 words in total)

Initially, the possibility of creating a text with two distinct sections was contemplated The first section would contain sentences with NTIs and the second would contain sentences with minimal NTIs However, the text was written in a natural language, which means that each segment contains different features of that language, some are short, some long, some have long noun phrases, some use referential pronouns etc Creation of the text envisaged above would have been difficult and would have led to an incoherent and unnatural-sounding text Consequently, it was decided that the NTIs would be scattered throughout the text, i e some segments would contain NTIs and some would not

In order to simplify the analysis process at a later stage, it was hoped that each segment would contain only one unique NTI However, this proved to be impractical The goal of adequate representation of all NTIs meant that some segments had to contain more than one indicator This was taken into account during the analysis process Where two or more NTIs occurred in a sentence, it was anticipated that Translog and CNA would provide evidence of which NTI, if any, required the most cognitive processing

For those NTIs where there are only two occurrences, it is acknowledged that this is not a very high representation Nonetheless, a trade-off was required between a high level of validity (i e a relatively "normal" text), and a high level of NTI representation We certainly could have created a list of unrelated sentences with a higher representation of each NTI, but it was felt that a "normal" text was preferable to an artificial one because this is what post-editors and translators (mostly) work with

5.3.7 Number of Subjects

The topic of subject selection has already been discussed However, one question that was not addressed is the ideal number of subjects for a study such as this Krings (2001) used 16 subjects in his core experimental group. Campbell (2000) talks about increasing the number of subjects so that closure of behaviour can be observed, but he does not offer advice on what that number should be.

An important objective for this study was to use professional, as opposed to semi-professional, translators and to replicate a normal work scenario. Therefore, the number of subjects used was dictated by the number of suitable subjects offered by IBM. An additional consideration was the time required for data analysis. Although this could not be determined in advance of the analysis task, it was reasonable to assume that the twelve subjects would generate a significant amount of data which then had to be analysed for the three types of post-editing effort mentioned previously (i.e. temporal, technical and cognitive).

5.3.8 The Translation Assignment

Krings (2001), Jääskeläinen (1987, 1990) and Tirkkonen-Condit (1989) all state that it is important to give subjects a translation assignment in translation process research. As can be seen from Appendix B, the post-editors and translators were given specific instructions on the translation assignment. Post-editors were given the following instructions:

"Do a **FULL post-edit** on the German target so that it meets the following criteria:

- Any non-sensical sentences or phrases are repaired.
- Any inaccuracies in the information are fixed.
- Any mis-translation, non-translation or inconsistent translation of terminology is rectified.
- The text is understandable and stylistically acceptable to a German native speaker who needs to understand the contents of the document.

NOTE: It is not necessary to change text that is accurate and acceptable just for the sake of improving its style.

Insofar as possible, you should treat this as a professional job and not as an academic exercise."

Translators were given the following instructions:

"**Translate** the English source into German so that it meets the following criteria:

You produce an equivalent, accurate version of the text in the target language (German), which is understandable and stylistically acceptable to a German native speaker.

Insofar as possible, you should treat this as a professional job and not as an academic exercise "

Although not stated in the written instructions, all subjects were informed that they should do one full pass over the text and that they were not expected to perform a proof-read and edit on the text after the first pass

5.3.9 Reference Works

Reference works were not made available to subjects because any pauses in text processing while reference works were being consulted would simply be recorded as a long pause by Translog and, without the use of video recording, it would have been impossible to accurately establish the reason for this long pause As already mentioned, specialised terms were coded in the MT dictionary and in Translog and this set up the expectation that subjects would have little need of looking up terms in hard-copy or online dictionaries During subject selection, one of the guiding principles was to ensure that subjects were experienced and competent enough so that they did not need to look-up general vocabulary in a reference book or online

5.3.10 The Pilot Experiment

As previously mentioned, pilot tests are one of the measures recommended for preserving measurement reliability A pilot run of the experiment was undertaken in order to learn about the potential pitfalls in conducting the experiment To gain the most advantage from the pilot test, as many parameters as possible were made consistent with the main experiment, i e the same technology was used (IBM's MT engine, Translog), the SL and TL were the same and the text was the same The main differences were the subjects themselves, their level of experience and the location of the experiment Six student translators, with German as their first language, were recruited Four were asked to complete the post-editing task and two were asked to translate Because time was limited, they were given one hour to complete as much of the text as possible The pilot experiment helped to refine the instructions for the main experiment and reassured the researcher that the technology was stable As we will see in the chapter on Data Analysis, the pilot data were useful for comparison with the professional subjects' data when an explanation was required for unexpected dictionary look-up behaviour and have been stored safely for potential future comparative research between semi-professional and professional post-editing processes

5.3.11 Measuring Cognitive Effort

The first step in measuring post-editing effort was to measure the cognitive effort for each segment. The measurement of cognitive effort centred around the use of Choice Network Analysis. A template was devised which captured the following data: the segment number; the source sentence; the NTI's (if any); the raw MT output; the output from the three translators – for reference purposes only - a table comparing the output of all post-editors with the MT system output and a section for commenting on the post-editing effort for that segment. A separate file was created for each segment and all CNA analyses are available in Appendix H. See Appendix I for the raw output from the MT system and Appendix J for all final post-edited and translated files.

The graphic representation of Choice Networks differs here from that of Campbell (1999, 2000a) and Campbell and Hale (1999). In those publications, CNA is presented as a map of different choices made by translators (see Chapter 4). This researcher found that such a map was difficult and very time-consuming to draw in an electronic document. In addition, it did not present a clear graphic representation of the choices made by post-editors. Therefore, an alternative presentation of the data was found (i.e. table format).

One of the difficulties encountered during the application of Choice Network Analysis was what level of granularity to include in each cell of the table. In other words, should each cell only contain one word or could it contain syntactic constituents, for example, Noun Phrase, Verb Phrase, Prepositional Phrase, or functional units such as subject, object, predicate or, indeed, clauses? Much has been written about the difficulty of identifying "translation units". Kenny (forthcoming) provides a breakdown of how translation units are viewed in product-, process-, NLP- and corpus-based translation studies. She refers to Zabalbeascoa (2000: 121) who maintains that the identification of translation units from a retrospective, descriptive point of view involves "first finding meaningful bitextual pairs, which means that the length and nature of each segment is determined by the type of solution, which provides evidence of the problem as the translator presumably saw it". This approach describes in essence the guiding principles used in this study for identifying units of analysis during Choice Network Analysis.

With sentences sometimes as long as 20 or more words, it became clear that one word per cell would result in very cumbersome tables. The researcher therefore let each segment dictate the granularity of each table. Where it was clear that no changes had been made by any post-editor to an entire clause or phrase, for example, that clause/phrase was housed in one cell (e.g. see the first cell in CNA Segment 112 as an example). Where, on the other hand, it was clear that changes had been made by multiple post-editors on a lower linguistic level, at the level of modifier, for example, one cell was dedicated to the modifier (see cell 5 in Segment 112) When a sentence was long, it was broken into two tables Clear presentation of the data and facilitation of the analysis were the guiding principles

A second problem arose when deciding how to present data where the post-editors had altered the syntax of the translation compared with the syntax of the MT output The solution was to keep the grammatical elements of the sentences aligned (for example, the verb in post-edited sentences was always aligned under the corresponding verb in the MT output), but to signify changes in syntax by sequentially numbering those elements where syntax had been altered (see Segment 7 as an example) By keeping the grammatical elements of each sentence aligned, it was then possible to highlight (using the **Bold** feature in Word) those parts of the sentence which had been the focus of post-editing activity

One of the problems that arose when applying CNA was how to differentiate between TT elements where there was a lot of variety in solutions (e g each post-editor rendered a different solution) and TT elements where, for example, all but one post-editor gave the same solution In CNA theory, the first case (nine different solutions) would indicate greater text difficulty than the latter Rather than accord equal status to choice nodes where only one post-editor had changed the MT output and those where more than one post-editor had made changes, it was decided to comment on *all* changes to the MT output in the appropriate section of the Choice Network Analysis template, but to highlight in bold only those changes made by two or more post-editors Although this is somewhat arbitrary, the strategy was to give more weight to changes made by a larger number of post-editors

Once the post-editing activity had been recorded in this way for each segment, a conclusion was drawn regarding the elements of the sentence that had been the focus of post-editing activity Care was taken here to take the specific NTIs in each sentence into consideration to see if they could be linked to post-editing effort

The second step in measuring cognitive effort involved the measurement of time spent pausing while post-editing These data were captured at the same time as data on technical and temporal effort and are described in more detail in the section on measuring technical and temporal post-editing effort

5.3.12 Measuring Technical and Temporal Post-editing Effort

The collation of data on technical and temporal post-editing effort involved a number of steps and sub-steps First it was necessary to capture the amount of time spent by each post-editor and translator on each segment Translog was used to capture these data and the log file was replayed during analysis of each segment One issue that arose here was how to correctly identify the boundaries between segments From the point of view of technical effort, this was easy the end of the segment came when a subject ceased to make changes to that segment and started to make changes to another one From a temporal and cognitive point of view, however, identifying segment boundaries was more challenging because it was not possible to determine exactly when a subject started processing a new segment It was decided that cursor movement would be used to identify segment boundaries When the cursor moved off the segment currently being edited and onto another segment, and if it did not move back onto the first segment, then a segment boundary was deemed to have been crossed This is, of course, problematic because we cannot be assured that the post-editor has stopped *thinking* about a particular segment when the cursor moves off that segment However, experience with Translog showed that once the cursor moved from one segment to the next, the post-editor quickly commenced editing the new segment and rarely went back to the previous one This suggests that cognitive processing of the new segment commenced at the point when the cursor was on the new segment, if not indeed before that The method has the added benefit of being systematically implementable, and applicable to all segments

When a segment boundary was identified using Translog, the log file was paused immediately and the time-stamp in the log file was recorded (see Appendix K for all linear repetition file data) The *total processing time* in seconds was recorded by subtracting the previous segment's final time-stamp from the current segment's time-stamp The number of source words was recorded for every segment To record the *processing speed* for each segment, the number of source words in the segment was divided by the *total processing time* The *processing speed* therefore gave a measurement which allowed us to make comparisons on temporal effort across segments, regardless of the number of source words and the existence, or not, of NTIs in the segment

The segment boundary was recorded in the Translog linear repetition file using the following end-of-segment marker [End Seg X], where "X" stands for the segment number The amount of processing time spent *pausing* was recorded by counting the number of seconds where no keyboarding or mouse activity occurred (Translog records this in the log file using the asterisk symbol) This gave the *total pause time* The *pause ratio* (i e the total pause time as a percentage of total processing time) was derived by dividing the total pause time by the total processing time This measurement indicates what percentage of time a subject spent pausing for each segment

The total processing time and processing speed were also calculated for each translator. The processing time was used as input for the calculation of *Relative Post-editing Effort (RPE)*. In keeping with Krings's (2001) suggestion, to calculate the RPE value for each segment, the average processing time for post-editing was divided by the average processing time for translation. This gave an indication of how translation effort compared with post-editing effort.

To record the technical effort, the Translog log file was replayed for each post-editor and the number of words (or parts of words) and punctuation marks inserted or deleted was recorded and then entered into the spreadsheet for each segment. The recording of partial deletions and insertions was considered to be important because such partial word manipulations are characteristic of post-editing activity (here I am drawing on what I have observed during this research) and not to record them would have produced an inaccurate picture of the full technical effort. In addition, any cut/copy and paste actions were recorded as well as search procedures in Translog's dictionary. The latter were classified into search procedures where the term was found and searches where the term was not found.

For many of the results reported in Chapter 6, the median value across post-editors is used rather than the mean. According to Cohen and Holliday (1982: 31), the mean is used when:

1. The scores in a distribution are more or less symmetrically grouped about a central point.
2. The research problems require a measure of central tendency that will also form the basis of other statistics (e.g. variability).
3. The research problem requires the combination of the mean with the means of other groups.

Alternatively, the median is used when:

1. The research problem calls for knowledge of the exact midpoint of a distribution.
2. Extreme scores could distort the mean.
3. Distributions are "oddly-shaped".

It was decided that the research problem and data analysed here were more in keeping with the latter. For example, while the measures for pause ratio were normally distributed, measures of relative post-editing effort, number of words inserted or deleted, etc., were not normally distributed. Therefore, the median was used.

Once all data were entered into a spreadsheet for each segment, those data were then transferred to an SPSS spreadsheet for inclusion in later statistical analyses. For an example of the data spreadsheets see Appendix D. Data analysis involved observing and reporting on temporal effort, e.g. total time required to complete the task, processing speed etc., technical effort, e.g. words inserted/deleted, cuts and pastes etc., and cognitive effort, e g pause ratio, Choice Network Analysis All of the above are reported in detail in Chapter 6

5.4 CONCLUSION

In this chapter we have presented the methodological framework for this research, both from a theoretical point of view, taking variables, operationalisation and validity into account, and from a practical point of view, taking technology, text type, data capture etc into account The decisions reported here were then put to practical use during data collation and analysis, the results of which are reported in Chapter 6

Chapter 6

6. DATA ANALYSIS

6.1 INTRODUCTION

The data on post-editing effort accumulated during this study will be presented and discussed in this Chapter We will first present and discuss data on temporal effort (Section 6 2) Next we will discuss data on technical effort (Section 6 3) and, finally, we will discuss data on cognitive effort (6 4) In Section 6 5 we will summarise our findings

6.2 TEMPORAL EFFORT

6.2.1 Time required for Post-Editing & Translation

All post-editors and translators were told that they had a maximum time of two hours to either post-edit or translate the text They were also told that they should complete a first pass only and that they were not expected to revise the text The total number of source words in the ST was 1,777 [59] As can be seen in Table 6 1, all post-editors completed the task within the two hours (with the exception of P6 who was just short of finishing by 16 words) and none of the translators completed the task within the allotted time-frame, although they all came close Figure 6 1 gives a graphic illustration of the data

Subject ID	Number of Source words processed	Time in minutes	Words per minute
P1	1777	64	27 77
P2	1777	92	19 32
P3	1777	101	17 59
P4	1777	123	14 45
P5	1777	79	22 49
P6	1761	119	14 80
P7	1777	100	17 77
P8	1777	118	15 06
P9	1777	104	17 09

[59] There were 165 segments in the ST A decision was made to finish the analysis at segment number 130 In creating the ST for analysis, four segments (2 and 4) were divided into two in order to spread NTIs across segments In addition, segment 41 was divided into three (41a, 41b and 41c) As a result, although the analysis finished at segment number 130, there was a total of 134 segments in the analysis The reasons for analysing only as far as segment number 130 were 1 The translators had not translated beyond this point so a comparison between translation and post editing effort was no longer possible 2 Subject P6 appears to have opted out of the research at this point His/her data become unusable after segment 130 because errors are introduced and not corrected 3 The first 134 segments accounted for 1 635 (or 92%) of the 1 777 words in the ST and all of the NTIs contained in segments 131 165 were already accounted for

Subject ID	Number of Source words processed	Time in minutes	Words per minute
T1	1623	120	13 52
T2	1585	120	13 20
T3	1635	120	13 63

Table 6 1 Time Required to Post-Edit and Translate ST

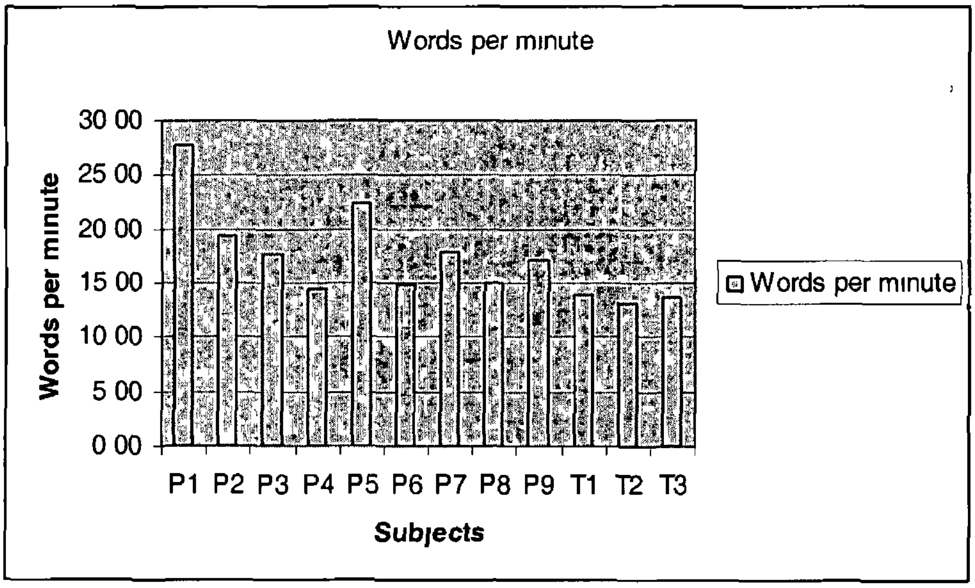


Figure 6 1 Graphic View of Words Per Minute Post-Editing vs Translating

The data show that subject P1 was by far the fastest post-editor, followed by P5 The speeds for P4, P6 and P8 were only slightly faster than for the three translators

It is interesting to note that the three translators' speeds were very similar (13 52, 13 20 and 13 63) This is reassuring as it suggests that the three translators' ability to comprehend the ST and produce a TT is similar We cannot, of course, make any claims about competence from the point of view of TT quality without performing an evaluation of the translation output, which is beyond the scope of this study Although the number of translation subjects is small, the similarity in their rate of words per minute is a positive factor - in particular because this measure is used in the calculation of *Relative Post-Editing Effort (RPE)*, which will be discussed later

Table 6 2 shows the median rate of words per minute for post-editing and translating The median rates show that post-editing was faster than translation

Subject Type	Median words per minute
Post-Editors	17 59
Translators	13 63

Table 6 2 Median Number of Words Per Minute Post-Editing vs Translation

The relatively high processing speeds for P1 and P5 warrant some discussion From Segment 50 onwards, P5's post-editing behaviour changed Several errors were introduced into the TT and were not corrected The following is an example of the typical output from P5 at this point in the text (taken from Segment 71)

> *Der Edfugt uhrt die Anfangs-End-Tags tags automatp inweise aare ein, so daß Dinge wie Listen automatbeischt werde enden*

There is no evidence to suggest that P5 had technological problems at this point in the text We can only speculate that P5 either became very tired or that s/he no longer wanted to act as a subject and decided to make his/her data invalid [60] The lack of care in post-editing would clearly have contributed to a higher processing speed When it became clear that P5's data were invalid from Segment 50 onwards, a decision had to be made whereby all of P5's data were eliminated from the study Unfortunately, this reduced the number of post-editors to eight

No such explanation is available for P1's high processing speeds We can only speculate that P1 can post-edit at a faster rate than the other subjects A cursory glance at P1's output suggests that the quality is at least satisfactory A more in-depth quality analysis is a separate study that could be carried out at a later date

The data presented here give credence to the claim that the post-editing of MT output is faster than the translation of a text However, this claim should be treated with caution because this study does not take into account the time required to revise the text to publishable quality Based on the study at hand, we can state that post-editing is faster than translation *as a first-pass exercise* We cannot state with absolute certainty that this is the case if revision to publishable quality is required That is another potential study for the future

To conclude this section, it is important to note that if P1 and P5's data are excluded from the calculation of words per minute, the median value changes only from 17 59 to 17 09 As already mentioned, P5's data were eliminated from the study because they were unusable The higher processing speeds for P1 do not alter the median value radically

[60] No indication to this effect was given One lesson that can be drawn from this is that retrospective interviews with playback would be useful in identifying such problems and would offer the subject an opportunity to explain his/her decision Also, the prospect of a retrospective interview might act as a disincentive for a subject to opt out half way through an experiment

Moreover, it is entirely feasible that some people will post-edit at a much faster rate than others, and P1's data reflect this. Given these two considerations, P1's data are included in all analyses.

6.2.2 Post-Editing Processing Speed $S_{(nti)}$ vs. $S_{(min\text{-}nti)}$

The *Processing Speed* is the total number of source words in each segment divided by the total processing time for that segment. It acts as an indicator of the speed with which each post-editor can process each segment. The processing speed for each segment was captured for all post-editors and all translators. The segment types were divided into two: segments containing negative translatability indicators - $S_{(nti)}$ - and segments containing minimal translatability indicators – $S_{(min\text{-}nti)}$. There were 103 segments of the type $S_{(nti)}$ and 31 of the type $S_{(min\text{-}nti)}$. The reason for including a higher number of the former is that we were testing a number of different NTIs and, therefore, a number of instantiations of each NTI was necessary in the ST. The function of the $S_{(min\text{-}nti)}$ segments was to provide data for comparison with $S_{(nti)}$ segments. The median value for both types of segment is shown in Table 6.3.

Segment Type	Std. Deviation	Median
Segment Type 1 - with NTIs	.74730	.3500
Segment Type 2 - minimal NTIs	.98882	.4350

Table 6.3: Median Processing Speed S(nti) vs. S(min-nti)

The median processing speed for segments containing negative translatability indicators is lower (.3500) than for segments containing minimal negative translatability indicators (.4350). We note that the standard deviation from the mean is high for both. A more detailed examination of processing speed for both segment types might help to explain this. An initial examination of processing speeds for segments of the type $S_{(nti)}$ revealed that there were some values that were considerably different from the median values. For example, P1 spent only one second on segment 67 and this led to a very high processing speed (16.00 words per second for a 16 word segment). Similarly, P4 spent only two seconds on segment 49, which had eight source words, leading to a processing speed of 4.00 words per second. In these cases, the subjects either did not make any edits, or the edits were made so quickly that the processing speed recorded was considerably higher than the median. The graph generated was somewhat skewed because of these very high single instances. Therefore, to facilitate interpretation of the data these outliers were removed for segment type $S_{(nti)}$ [61] Figure 6 2 shows the pattern for processing speeds for segment type $S_{(nti)}$ [62]

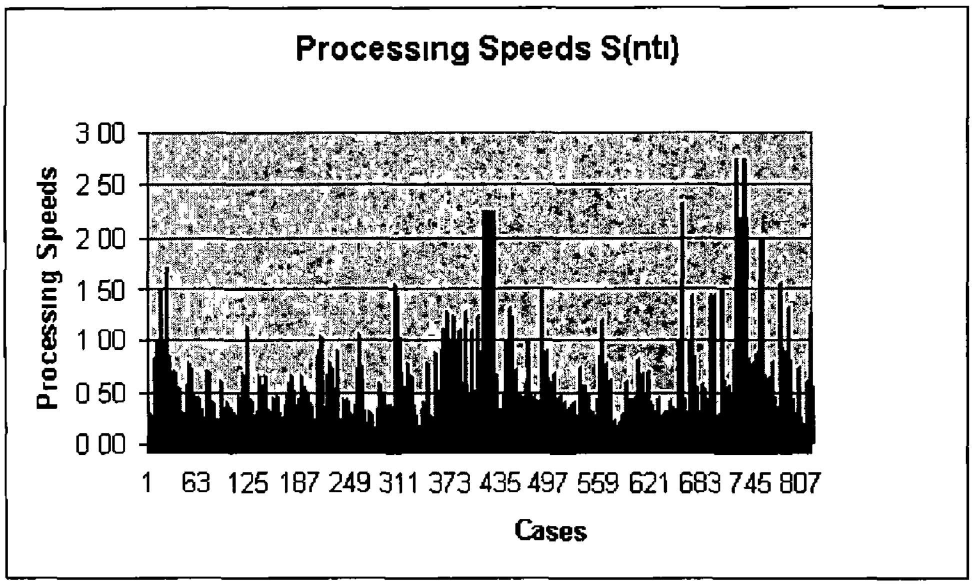


Figure 6 2 Processing Speeds for S(nti)

While the majority of cases fall below the 0 50 mark, there are a number of cases that range from 0 50 to 2 75 This explains why the original standard deviation for $S_{(nti)}$ was as high as 74730 Note that when the five high-value cases mentioned above are removed from the list of values, the standard deviation is reduced from 74730 to 36362 We can deduce from these data that the processing speeds for $S_{(nti)}$ segment types can vary considerably

For segments of type $S_{(min\ nti)}$, there were also some outliers in the data [63] Figure 6 3 shows the processing speeds for $S_{(min\ nti)}$ segments with these outliers removed [64]

[61] Five individual values were removed, all above the value 3 00

[62] Processing Speeds' on the Y axis are measured in words per second "Cases" on the X axis refer to the number of data points recorded i e 103 $S_{(nti)}$ segments multiplied by eight post editors minus the five outlier values which is equal to 819 data points SPSS generated a graph with data points displayed in increments of 62

[63] In segment 63, P1's speed was 8 00 and in segment 125, P3 and P4 also scored 8 00 while P4 scored 4 00 in segment 49 and P1 scored 5 00 in segment 10

[64] Again "Processing Speeds" are measured in words per second and "Cases" represent the number of data points (31 segments by eight post editors minus five outliers (= 243) The data are displayed in increments of 18

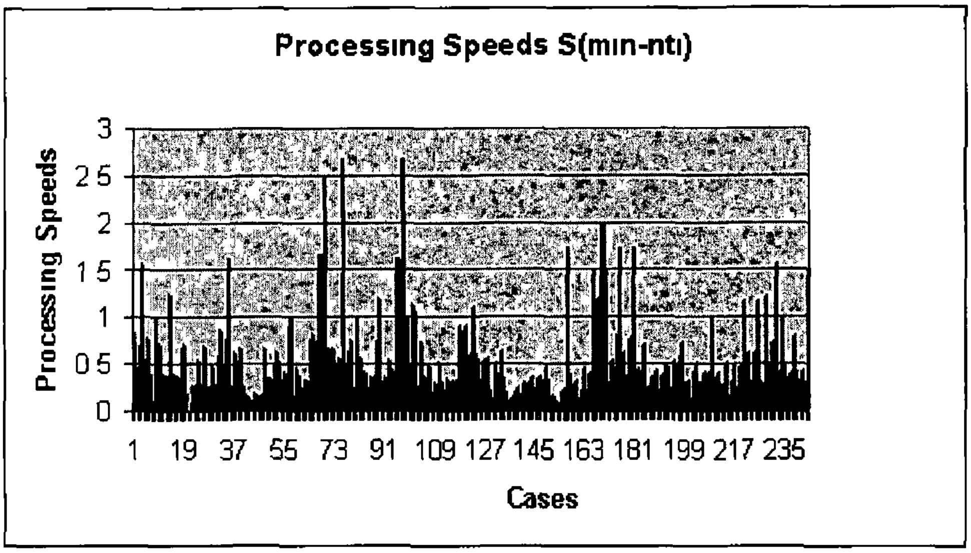


Figure 6 3 Processing Speeds for S(min-nti)

As with the processing speeds for $S_{(nti)}$ segments, a large proportion of cases have values below 0 50, but there are also a number of cases with values between 0 50 and 2 75, indicating that here too the processing speeds can vary considerably for this segment type With the five outliers removed, the standard deviation is reduced from 98882 to 44626

There is an obvious question to ask here Given that there are 819 records of processing speed for the $S_{(nti)}$ segment type and only 243 records for the $S_{(min\ nti)}$ type, can we make valid comparisons between the two sets of data? In order to answer this, SPSS was used to randomly select 243 records of processing speed for the $S_{(nti)}$ segment type and the median processing speed and standard deviation were calculated for this sample The value returned for the median processing speed was 3500, which is identical to the value for the 819 cases The standard deviation was 1 08987, which is higher than the 74730 calculated for the 819 cases and is probably influenced by the fact that two of the outlying values were selected in the random selection (1 case of 16 00 and 1 case of 3 00) This result means that we can make comparisons between the two sets of data without being concerned about the difference in numbers of actual cases

In summary then, the median processing speeds suggest that segments that contain minimal negative translatability indicators can be processed faster than those containing NTIs The H_0 states that there is no difference between the processing speeds for the two segment types A Mann-Whitney U test returns a significance value of p = 000 [65] Since p<0 05 we can reject the null hypothesis and say that the differences between the processing speeds for the two segment types are statistically significant Although it appears that the $S_{(min\ nti)}$ segment type can be processed faster than the $S_{(nti)}$ segment type, we should keep in

[65] The Mann Whitney U test was deemed to be most appropriate in this case as it is the best known non parametric statistical significance test

mind that the former is not free of post-editing. The data suggest that even if all NTIs are removed from a text using a CL checker, post-editing may still be required (See Section 6.4 for more discussion on this).

Let us now examine the NTIs that occur in segments with low processing speeds with the aim of establishing whether or not specific NTIs have a high representation in these segments. At this stage we will not examine the effect of the NTIs in individual segments as this will be done later in our section on Cognitive Effort. Nonetheless, by looking at segments with low processing speeds we might at least get an indication of which NTIs are most problematic for post-editing. One of the difficulties here is defining "low" processing speed. We know that .3500 words per second is the median processing speed for $S_{(nti)}$ segments. Therefore, let us use this as an arbitrary cut-off point. Table 6.4 shows the NTIs that occur in segments with processing speeds below this value.

NTI	Total Number in ST	Number in segments with low processing speed	% of this NTI occurring in segments with low processing speed
Gerund	24	12	50%
Proper Noun	25	20	80%
Abbreviation	18	7	39%
Punctuation	17	9	53%
Use of slash	3	3	100%
Ungrammatical construct	3	2	67%
Post-modifying adjectival phrase	2	2	100%
Use of (s) for plural	3	3	100%
Non-finite verb	4	2	50%
Slang	2	1	50%
Misspelling	3	2	66%
Personal Pronoun	16	2	13%
Not a full syntactic unit	5	5	100%
Long Noun Phrase	10	6	60%
Ambiguous scope in coordination	3	1	33%
Ellipsis	9	1	11%
Missing relative pronoun "that"	3	1	33%
Passive Voice	5	1	20%
Contraction	2	1	50%
Demonstrative Pronoun	7	1	14%
Short segment	6	3	50%
Minimal NTIs	31	10	32%

Table 6.4: NTIs in Segments with Low Processing Speeds

The data ın Table 6 4 ındicate what percentage of ındıvıdual NTIs occur ın segments wıth processıng speeds below the medıan value of 35 words per second We observe that some NTIs have a low representatıon (albeıt based on very few ınstances to start wıth), e g abbrevıatıon, ellıpsıs, passıve voıce and demonstratıve pronoun Others have a hıgh representatıon, e g ungrammatıcal construct and use of (s) as a plural marker Although thıs gıves us an ındıcatıon of what NTIs may have affected processıng speed, ıt does not demonstrate whether or not the NTIs ın questıon dıd, ın fact, cause problems for post-edıtors That questıon wıll be answered ın the sectıon on Cognıtıve Effort

6.2.3 Relative Post-Editing Effort

As ındıcated ın Chapter 3, Krıngs (2001) defınes *relatıve post-edıtıng effort* as the tıme requıred to machıne translate and post-edıt a text dıvıded by that requıred for translatıon A value between zero and one means that MT plus post-edıtıng effort was less than translatıon effort A value of one means that translatıon and post-edıtıng effort are equal and a value greater than one means that MT plus post-edıtıng ıs more tıme-consumıng than translatıon

Krıngs correctly ımplıes that a comparıson of translatıon and post-edıtıng effort should ıdeally take ınto account the tıme spent preparıng the fıle for machıne translatıon Thıs mıght ınclude controlled language checkıng and edıtıng, termınology codıng and fıle format conversıon, where necessary As already mentıoned, 79 terms were coded ın the MT dıctıonary and thıs dıd not take more than one hour altogether The tıme requıred for machıne translatıon (carrıed out ın-house by IBM) amounted to only a few seconds Sınce ıt would have been dıffıcult to accurately spread the short tıme requıred for dıctıonary codıng and machıne translatıon over the tıme requıred for post-edıtıng all segments, both tasks have been elımınated from the calculatıon Here we recall Ryan's (1988 131) contentıon, mentıoned ın Chapter 3, that the speed wıth whıch computers can translate these days renders the tıme requıred for MT ınsıgnıfıcant and that we should ınstead focus on the *effectıveness* of the post-edıtıng process The tıme taken to edıt the ST ıs not ıncluded eıther because the edıtıng process was an unusual one ın that some NTIs were elımınated and others were delıberately ıntroduced Constant trackıng and recordıng of the edıtıng process was necessary for methodologıcal reasons To take the edıtıng tıme ınto account, then, would gıve a false ımpressıon of the tıme normally requıred for thıs task To conclude, the measurement of relatıve post-edıtıng effort for thıs study ınvolves a sımple comparıson of post-edıtıng effort and translatıon effort and ıs calculated for each segment by dıvıdıng the average processıng tıme for post-edıtıng by the average processıng tıme for translatıon Table 6 5 ıllustrates the data collectıon method usıng segment 118 as an example

Subject	No of source words	Start Time mm ss	End Time mm ss	Total Processing Time (sec)	Processing Speed
Subject P1	14	51 42	51 51	9	1 56
Subject P2	14	15 25	15 48	23	0 61
Subject P3	14	05 31	05 59	28	0 50
Subject P4	14	37 03	37 36	33	0 42
Subject P6	14	39 09	39 24	15	0 93
Subject P7	14	06 42	06 58	16	0 88
Subject P8	14	39 52	40 09	17	0 82
Subject P9	14	10 13	10 38	25	0 56
Subject T1	14	46 56	47 12	16	0 88
Subject T2	14	51 31	52 10	39	0 36
Subject T3	14	49 44	50 08	24	0 58
Average for Post-editing				20 75	0 79
Median for Post-editing				20	0 72
Std Deviation for Post-editing				7 36	0 34
Average for Translation				26 33	0 61
Median for Translation				24	0 58
Std Deviation for Translation				9 53	0 21
Relative Post-Editing Effort	0 79				

Table 6 5 Example of How Relative Post-Editing Effort is Calculated

Let us first take a look at the Relative Post-Editing Effort for all segments Figure 6 4 presents a graphic representation of these data

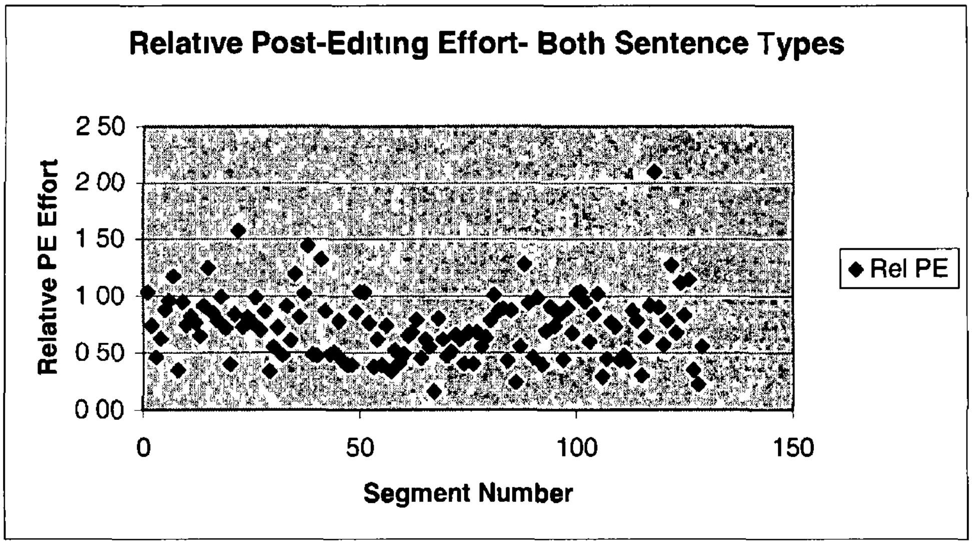


Figure 6 4 Relative Post-Editing Effort for Both Segment Types

For the majority of segments, the Relative Post-Editing Effort lies below 1 (19 out of 130, or 15% are above 1) [66] According to Krings, this implies that the post-editing effort for these segments was lower than the translation effort for the same segments This confirms our finding with regard to post-editing speed and translation speed, i e that post-editing was by and large faster than translation It is noteworthy that of the segments that lie below 1 00, 76 segments (or 58%) lie between 0 50 and 1 00

While Figure 6 4 portrays a picture of the overall Relative Post-Editing Effort, it does not differentiate between the two segment types $S_{(nti)}$ and $S_{(min\ nti)}$ Figure 6 5 provides the values on the basis of segment types

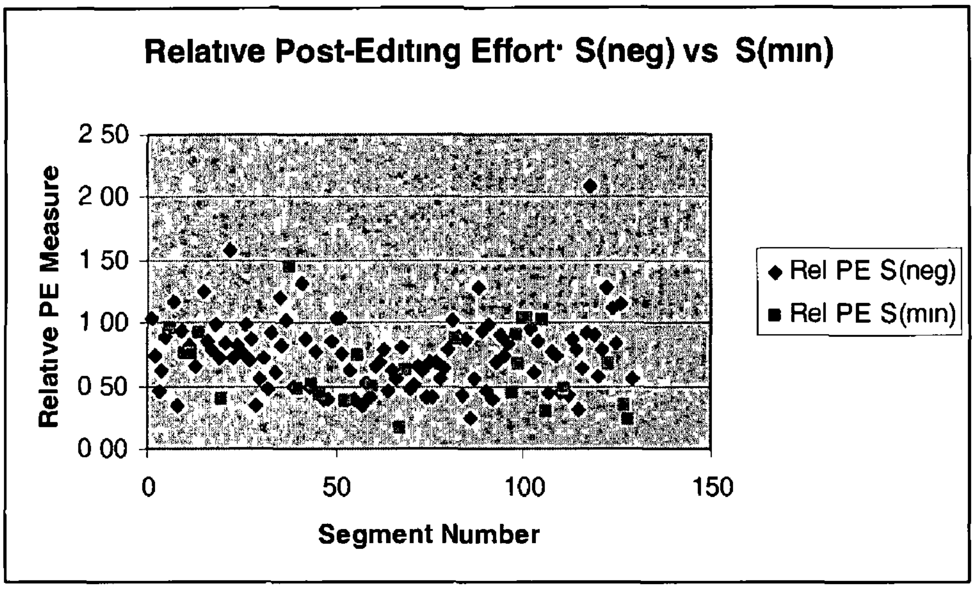


Figure 6 5 Relative Post-Editing Effort S(nti) vs S(min-nti)

At a first glance, the general pattern for Relative Post-Editing Effort for the $S_{(min\ nti)}$ segment type appears to be similar to that of $S_{(nti)}$ The majority of measures are below 1 00 (22 out of 27, or 81%), but 15% (4 segments out of 27) are above 1 00 This suggests that the RPE for most of the segments classified as $S_{(min\ nti)}$ is lower than the translation effort Nonetheless, for 15% of the $S_{(min\ nti)}$ segments, post-editing effort was greater than translation effort On the other hand, we should also note that the proportion of $S_{(min\ nti)}$ segments that are under 0 5 is greater, at 41% (or 11 out of 27), than the proportion of $S_{(nti)}$ segments under 0 5 (19% or 20 out of 103) This suggests that for a higher proportion of $S_{(min\ nti)}$ segments the RPE was considerably lower than translation effort Thus we can hypothesize that although some $S_{(min\ nti)}$ segment types might incur a high Relative Post-Editing Effort, a good proportion will incur a low RPE suggesting in turn that the removal of NTIs may indeed reduce RPE overall

[66] 130 segments are considered here (up to segment number 126, plus the four segments that were created when original segments were divided into smaller ones, e g 2a and 2b) instead of the 134 analysed in total because the translators translated only as far as segment number 126 within the allocated time which means that it is not possible to calculate the Relative Post Editing Effort for the remaining segments

What general conclusions can we draw from an analysis of these measures? The H_0 in this case states that there is no difference in RPE between the two segment types A Mann-Whitney U test returns a p value of 0 165 Given that p>0 05 in this case, we cannot reject the null hypothesis Therefore, there does not appear to be a significant difference between Relative Post-Editing Effort for the two segment types

Thus far we have presented a general picture of RPE for the two segment types and we have seen that both low and high RPE values have been recorded for both segment types In the next section we identify features in the segments that contribute to a high RPE

For the purposes of analysis, we will define "high RPE" as an RPE value above 0 90 This is an arbitrary cut-off point, but since 0 90 means that the RPE is very close to the translation effort, the linguistic features of these segments might be worthy of closer investigation 32 segments are included in the set of segments with an RPE value equal to or greater than 0 90 (i e 25% of the total number of segments under consideration here)

Appendix L gives a list of the segments containing RPE values above 0 90 Of the 32 segments, 7 are classified as $S_{(min\ nti)}$ The NTIs contained in each of these 32 segments are also listed in the Appendix Table 6 6 gives a list of the NTIs contained in segments with high RPE values

NTI	Total Number in ST	Number in segments with high RPE Values	% of this NTI occurring in segments with high RPE values
Gerund	24	7	29%
Proper Noun	25	3	12%
Ungrammatical construct	3	2	67%
Use of (s) for plural	3	1	33%
Non-finite verb	4	2	50%
Not a full syntactic unit	5	1	20%
Long Noun Phrase	10	2	20%
Ellipsis	9	3	33%
Passive Voice	5	1	20%
Short segment	6	2	33%
Problematic Punctuation	17	4	24%
Multiple Coordinators	2	1	50%
Long Sentence	3	1	33%
Use of Parentheses	3	1	33%

Table 6 6 NTIs in Segments with High RPE Values

In addition to the NTIs in this list, a number of problems in these segments were caused by linguistic items that are not classified as NTIs These items are discussed in detail in Appendix H An overview of the types of problem concerned is included here

Terminological Problems

"Terminological Problems" cover specialised terms that ought to have been coded in the MT dictionary, but were not To recap, 79 terms were coded, but it is clear that more terms ought to have been coded as some poor translations of terms contributed to post-editing effort For example, terms such as *toolbar button* (Segment 16), *C drive* (Segment 46), *title tag* (Segment 88), *DBODY tags* (Segment 91), *P element* (Segment 97) and *insert markup dialog* (Segment 102) all contributed to problems in the segments with RPE values above 0 90

Polysemy

The polysemous nature of words such as *letter* and *ordered* (both Segment 35) contributed to post-editing effort

Verbs

The CNA evidence suggests that the MT output for verbs was frequently changed by post-editors (there were changes made to verbs in 37 segments) In the segments with high RPE, examples include *is* (Segment 85), *opens* (Segment 33), *move* and *select* (both Segment 91), *indicates* (Segment 97) and *provides* (Segment 123)

Formulaic Expressions

Expressions that could be described as "formulaic" also seem to cause recurring problems for post-editors For example, *See X for more information*, where "X" refers to another part of the document, is problematic (Segments 36 and 121) In addition, phrases that are common in this text type also caused problems for post-editors, e g *From Windows Explorer* (Segment 47) and *From within the Project Folder* (Segment 12) It is possible that German translators expect such phrases to be expressed according to a familiar formula in German (e g *Weitere Informationen finden Sie unter* or *Im Windows Explorer wahlen Sie*) and that if the MT output does not conform to this, they feel obliged to rectify the "error"

Just as an examination of NTIs in segments with high RPE values will add to our understanding of which NTIs generate most post-editing effort, an examination of NTIs in segments with low RPE values will add to our understanding of NTIs that do not generate much post-editing effort Again we are faced with the difficulty of deciding what "low" RPE actually means Since 92 out of 130 (or 71%) of segments for which RPE values are available lie between 0 50 and 1, we will classify anything lower than 0 50 as having a "low RPE value" (although this is another arbitrary threshold) The 34 segments with low RPE values (26% of the total under consideration) are contained in Appendix M

Of the 34 segments with RPE values below 0 50, 10 are classified as $S_{(\text{min nti})}$ Eight of the segments contain a proper noun The CNA shows that, in many cases, a proper noun does not cause problems for MT, as long as the MT system has a facility for recognising the proper noun as such Table 6 7 gives a list of the NTIs that occur in segments with a low RPE value

NTI	**Total Number in ST**	**Number in segments with low RPE Values**	**% of this NTI occurring in segments with low RPE values**
Proper Noun	25	8	32%
Not a full syntactic unit	5	3	60%
Ellipsis	9	2	22%
Missing "in order to"	5	4	80%
Missing relative pronoun "that"	3	3	100%
Personal Pronoun	16	4	25%
Problematic Punctuation	17	4	24%

Table 6 7 NTIs in Segments with Low RPE Values

To conclude on relative post-editing effort, there are a number of NTIs that have a high representation in segments with both high RPE and low RPE that are worth drawing attention to The NTIs that occur in segments with high RPE and that account for more than 50% of the occurrences of these NTIs in the ST are non-finite verb, ungrammatical construct and the use of parentheses For segments with low RPE, the NTIs with a high representation are missing "in order to", missing relative pronoun "that" and "not a full syntactic unit" From these data we can suggest that, in the former case, these NTIs are likely to contribute to a high RPE In the latter case, the occurrence of these NTIs does not appear to contribute to post-editing effort We will comment further on these NTIs in the section on Choice Network Analysis

6.2.3.1 Conclusions on Temporal Effort

The conclusions of our analysis of temporal post-editing effort are as follows

- The post-editing task was completed faster than the translation task
- The median processing speeds for $S_{(\text{min nti})}$ segments were higher than $S_{(\text{nti})}$ segments and the differences were statistically significant

- However, individual processing speeds for $S_{(min\ nti)}$ segments were not consistently higher than the speeds for $S_{(nti)}$ segments $S_{(min\ nti)}$ segments can sometimes have lower processing speeds than $S_{(nti)}$ segments

- The Relative Post-Editing Effort for the majority of segments is lower than translation effort

- Compared with $S_{(nti)}$ segments, a higher proportion of $S_{(min\ nti)}$ segments has an RPE value below 0 50, but we cannot state that there are statistically significant differences between the two segment types

- An analysis of NTIs shows that some NTIs lead to a higher RPE and a lower processing speed than others, suggesting that they are problematic for post-editing For example, the NTIs “ungrammatical construct” and “use of (s) as a plural marker” have a high occurrence in segments with low processing speeds Along with the NTIs “non-finite verb” and “use of parentheses”, the NTI “ungrammatical construct” also has a high occurrence in segments with a high RPE value NTIs that occur frequently in segments with low RPE values include “not a full syntactic unit”, “missing in order to” and “missing relative pronoun that” We cannot draw conclusions on NTIs here, but later on will discuss correlations between these findings and data gleaned from Choice Network Analysis

6.3 TECHNICAL EFFORT

Technical effort is gauged here by measuring both keyboarding and dictionary look-up activity For keyboarding, this meant using Translog to count the number of deletions, insertions, cuts and pastes When a word, or part of a word, was deleted it was counted as one deletion When a word or part of a word was inserted, this was counted as one insertion and so on Since Translog does not have a facility to report the number of words or parts of words deleted or inserted, this was done manually The second aspect of technical effort is dictionary look-up The number of look-ups was also recorded using Translog The use of the dictionary look-up facility will be commented on in more detail below

6.3.1 Keyboarding

Table 6 8 shows the median values for deletions, insertions, cuts and pastes according to segment-type

Segment Type	Median Deletions	Median Insertions	Median Cuts	Median Pastes
$S_{(nti)}$	4 00	4 00	00	00

$S_{(min\ nti)}$	3 00	3 00	00	00

Table 6 8 Deletions, Insertions, Cuts and Pastes by Segment Type

The median values for deletions and insertions are higher for $S_{(nti)}$ segments than for $S_{(min\ nti)}$ The standard deviation for deletions for $S_{(nti)}$ segments is 4 743 and 3 755 for $S_{(min\ nti)}$ segments For insertions, the standard deviations are 4 860 ($S_{(nti)}$) and 4 106 ($S_{(min\ nti)}$) respectively Cutting and pasting occurred so rarely for both segment types that the median values are 0 00 For $S_{(nti)}$ segments, 41 segments contained one cut action and 39 contained one paste action while three segments contained two cut actions and one segment contained two paste actions For $S_{(min\ nti)}$ segments, only eight segments contained one cut action and nine contained one paste action while no segments contained more than one cut or paste action

Our data suggest a difference, albeit a small one, in the technical effort required for post-editing $S_{(min\ nti)}$ and $S_{(nti)}$ segment types The significance value returned by a Mann Whitney test is p = 0 003 for deletions and p = 0 000 for insertions We can therefore reject the null hypothesis in this case and say that there are significant differences between the two segment types On the other hand, the p values for cutting and pasting are 0 173 and 0 418 respectively In this case we cannot reasonably reject the null hypothesis

The number of incidences of cutting and pasting is too small to allow us to make any generalisations about the differences between the two segment types Nonetheless, we can observe that cutting and pasting is rare in the post-editing activity observed in this study While analysing the Translog files, the researcher frequently observed that a word that already existed in the MT output would be duplicated by the post-editor and the original, identical word would be deleted In some cases, entire clauses were duplicated, character by character Therefore, the low occurrence of cutting and pasting is not due to the fact that little or no text could have been recycled How can we explain this behaviour then? We could suggest that duplicating a word by typing it is perhaps less labour-intensive and time-consuming than selecting it, cutting and pasting it elsewhere in a segment However, this does not explain why sometimes entire clauses are reproduced – the cutting and pasting of an entire clause would involve less keyboarding than retyping the entire clause Perhaps, as previously mentioned, the post-editor is working with certain cognitive units so that s/he ascertains that something needs to be edited and s/he reproduces the unit of translation without noticing that some elements of that unit already exist in the MT output and could be re-used? An additional explanation might be that the subjects in this study were simply in need of some training on how to recycle words more effectively for post-editing

We cannot arrive at any final conclusions here regarding the rare occurrence of cutting and pasting for these post-editors. However, we can make some recommendations for the future. Firstly, the data present an argument in favour of training people specifically in the task of post-editing. One of the aims of this training would be to reduce unnecessary keyboarding activity by post-editors. Secondly, and as some authors have suggested (Povlsen and Bech 2002, Allen and Hogan 2000, Knight and Chander 1994), there appears to be a strong case for developing a post-editing tool that would make it easier for post-editors to recycle elements of a segment without requiring intensive keyboarding or mouse movement. Post-editing trends for specific target languages would have to be coded into such a tool. Vasconcellos (1986a, 1986b) has commented on this for the English-Spanish language pair and a similar analysis could be done for other language pairs and specialised domains.

6.3.2 Use of Dictionary

As previously mentioned, Translog has a facility which allows the user to record source and target terms in the dictionary. The translator can then search the dictionary while translating. It was expected that dictionary look-up could be used as a measure of technical effort.[67]

As indicated above, some specialised terms from the ST were recorded in the dictionary along with their target terms. Before commencing their tasks, the post-editors and translators were informed that some terms were included in Translog's built-in dictionary. They were given a demonstration of how to perform a search in the dictionary and they were also given specific instructions on how to practise this search, as well as time to do a practice run before commencing their tasks. All subjects reported that they were familiar and comfortable with the dictionary look-up facility before commencing their tasks.

The Translog files report that very few subjects made use of the dictionary look-up facility. When they did use it, they frequently used it incorrectly so that they did not get a response from the dictionary. Table 6.9 summarises the dictionary look-up activity.

Subject	Successful Dictionary Lookups	Unsuccessful Dictionary Lookups
P1	0	0
P2	0	1
P3	0	5
P4	0	0

[67] It was also anticipated that dictionary look-up could be used as an indicator of cognitive effort since searching for a specific term would indicate that the post-editor had a difficulty with that term.

Subject	Successful Dictionary Lookups	Unsuccessful Dictionary Lookups
P5	0	1
P6	1	0
P7	0	1
P8	0	5
P9	0	1

Table 6 9 Dictionary Look-Up Activity

P1 and P4 did not attempt any searches in the dictionary P6 was the only subject to perform a successful search, i e the subject searched for the word *entity* in segment 9 and the dictionary returned a suggestion for that term While P3 and P8 both performed five searches, they were unsuccessful for two different reasons P3 did not find any of the terms in the dictionary (e g *feetures* and *nifty* in segment 22) while P8 did not perform the search correctly - s/he forgot to select the source term first using the mouse or cursor keys The same problem was encountered by P2, P5, P7 and P9

We can conclude that some subjects were not as familiar with the dictionary look-up facility as they had reported to be We can also surmise that many subjects did not feel that it was necessary to use the dictionary for support and that this is perhaps due to their experience as translators in this domain and with this text type In addition, the subjects normally work with the IBM translation memory and term management tool and are used to having terms suggested to them on-screen This may also have influenced their dictionary look-up behaviour Alternatively, we can suggest that when they did not get the expected answer from the dictionary on the first few attempts, they gave up on the idea of using it

Given the results from the dictionary look-up analysis, we thought it useful to conduct a comparison with the dictionary look-up activity of translation students who participated in the pilot project for this study This comparison reveals some differences between the two types of user (termed professional and non-professional, following Jaaskelainen 1987 and 1990) Table 6 10 shows the data for the six non-professional translators

Subject	Successful Dictionary Lookups	Unsuccessful Dictionary Lookups
A	0	0
B	3	13
C	12	30
D	1	1
E	8	19
F	6	18

Table 6 10 Dictionary Look-Up Activity for Non-Professional Subjects

With the exception of Subject A, all subjects used the dictionary look-up facility in Translog more frequently than the post-editors in the main study When the non-professional subjects performed a dictionary look-up they had a successful result more often than the professional subjects However, they also had unsuccessful results and, despite the fact that they seemed to have mastered the dictionary look-up facility in Translog, the unsuccessful results all resulted from an incorrect look-up procedure rather than the fact that the word was not contained in the dictionary While subject A did not use the dictionary at all, subject C used it frequently In fact, subject C seemed very reliant on the dictionary, searching for words such as *specify* and *text* as well as specialised words such as *document type definition* If subject C did not find a term or word in the dictionary, the log file shows that s/he would sometimes perform a search for the same word again [68]

In general, the non-professional subjects were much more reliant on the Translog dictionary facility and their log files show that they had more success with their searches At the same time, they also had more unsuccessful searches than the professional subjects because they were more reliant on the dictionary The data on dictionary look-up demonstrate that Translog is perhaps in need of a more user-friendly look-up facility Due to the lack of use of the dictionary look-up facility by the professional subjects, we cannot use it as an indicator of technical or cognitive effort

6.3.3 Conclusions on Technical Effort

- $S_{(min\ nti)}$ segments require significantly fewer deletions *and insertions* than $S_{(nti)}$ segments This result on technical post-editing effort adds to the evidence presented above on temporal post-editing effort and further supports the claim that the elimination of NTIs from a segment can reduce post-editing effort
- Cutting and pasting is a very rare activity for both segment types
- Dictionary searches were uncommon during this study When they were carried out, the search facility was frequently used *incorrectly* This suggests that either the post-editors were so experienced that they did not need to perform searches in the dictionary and/or that they were not comfortable with the dictionary facility in Translog As a result, it was not possible to use dictionary searches as indicators of technical or cognitive effort

[68] Livbjerg and Mees (2003 123) report on research results which suggest that student translators overuse dictionaries and focus too narrowly on lexical units at the expense of other units, such as context

6.4 COGNITIVE EFFORT

Our analysis of cognitive effort will focus on pauses and Choice Networks In section 6 4 1 we will discuss pauses and in section 6 4 2 we will turn our attention to Choice Network Analysis

6.4.1 Pauses as indicators of cognitive effort

In Chapter 4, we discussed the merits of pauses as indicators of cognitive effort and we observed that there is general agreement that pauses are indicators of such effort The objective in this study was to examine pause behaviour in order to gain insight into cognitive effort To this end the total pause time was extracted from the Translog file for each segment and each subject, and was subsequently recorded in an Excel spreadsheet The proportion of processing time spent "pausing" was then calculated This measure was labelled the "pause ratio" and is expressed as a percentage of the total processing time For example, for segment 8, the total processing time for subject P6 was 27 seconds Of this time, 14 12 seconds were spent pausing, i e no keyboarding or mouse activity took place during that time The pause ratio for this subject was then 52% Pause ratio trends for both segment types are discussed in more detail below

Initially pause analysis also involved comparing pause behaviour across all post-editors for each segment By analysing the post-editing activity after each pause of one second or longer, it was hoped that correlations between pause duration and position on the one hand, and NTIs on the other could be identified However, it soon became clear that this type of pause analysis would not reveal generalisable results because we could not state with confidence that a post-editor was focusing on a problem caused by a specific NTI during a pause Pauses were frequently followed by cursor movements rather than text editing An example of a pause analysis for segment 2a is included in Appendix N Here we can see, for example, that P7 pauses for 15 32 seconds and then presses the down arrow to move on to the next segment During this pause, P7 may have been thinking about the text located at the cursor point However, it is also feasible that P7 was thinking about another problem elsewhere in the text or that s/he was thinking about something entirely different during this pause Jakobsen (2005 112) refers to the "chunking" of language production as "cognitive rhythm" and suggests that the cognitive rhythm "is no doubt sometimes affected by factors that are not directly related to the cognitive processes involved in the production of the target text" (ibid) The pressing of the down arrow by P7 after 15 seconds tells us nothing about what was going on in the post-editor's mind during this pause Similarly, P6 pauses on three occasions for 14, 24 and 17 seconds and these pauses are punctuated by the use of the Control Key and the right or left arrow While shorter pauses were also frequently followed by cursor movements, they were often followed by text edits too For example, P6 pauses for one second and then types *leistungs* We cannot state for certain that the decision to type *leistungs* was made during the one-second pause that precedes this text edit Jakobsen (2005 113) also draws attention to this problem when he asks if cognition is necessarily restricted to occurring during pauses

A visual analysis of each pause and the subsequent post-editing activity for all post-editors and all segments up to segment 50 suggested that correlations between the duration and placement of pauses and NTIs were difficult to establish for the reasons stated above In addition, there was no discernible difference in the pause behaviour for the two segment types The conclusion we drew from this was that a detailed analysis of pause duration and placement did not contribute significantly to the findings already offered by Choice Network Analysis and so no further analysis of pause behaviour was carried out after segment 50

6.4.1.1 PAUSE RATIO $S_{(NTI)}$ VS. $S_{(MIN\text{-}NTI)}$

The expectation for pause ratios was that the more a segment required post-editing, the lower the pause ratio would be because more processing time was taken up with keyboarding than with pausing To take this expectation one step further, $S_{(nti)}$ segment types might be expected to have lower pause ratios than $S_{(min\ nti)}$ types because the former would require more post-editing effort Figure 6 6 shows the pause ratio values for the segment type $S_{(nti)}$ [69]

[69] For Figures 6 6 and 6 7 "Case Number" represents the data point allocated to pause ratio values in SPSS The total number of data points amounted to 1072 (i e eight post editors by 134 segments)

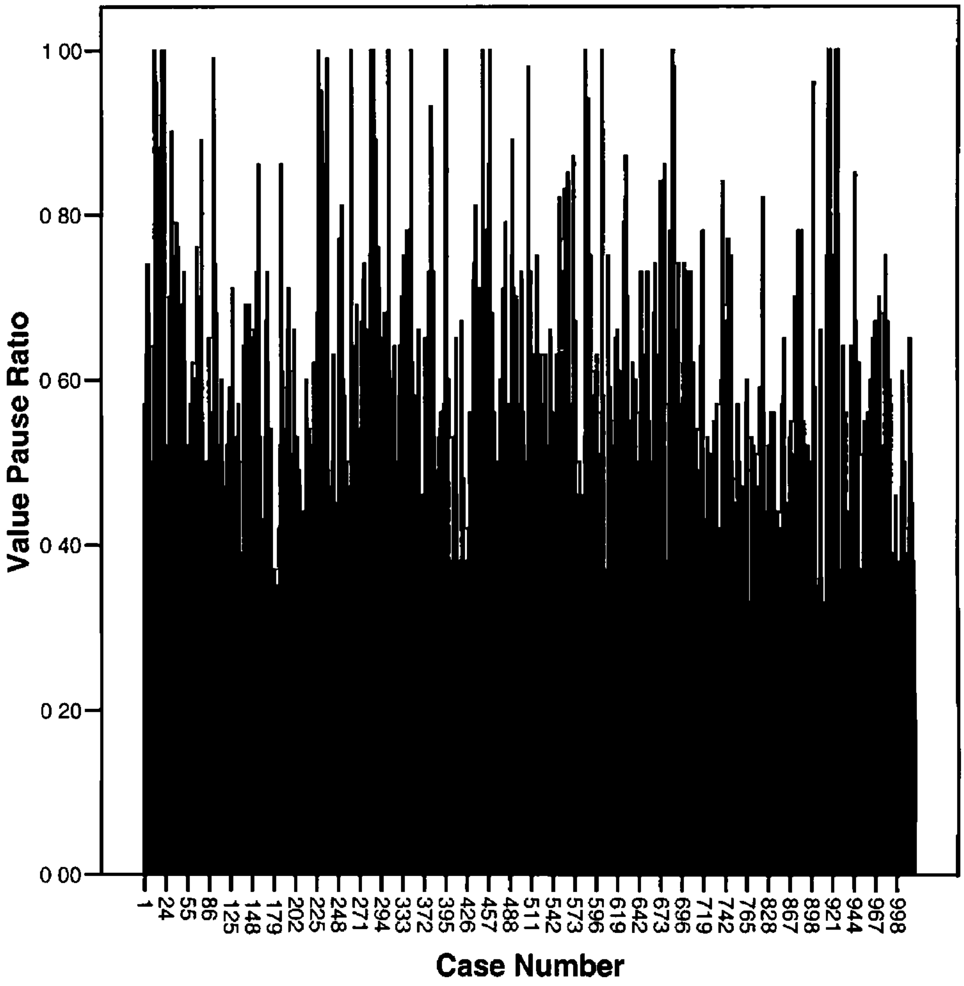


Figure 6 6 Pause Ratios for $S_{(nti)}$

Figure 6 7 shows the pause ratio values for the segment type $S_{(min\ nti)}$

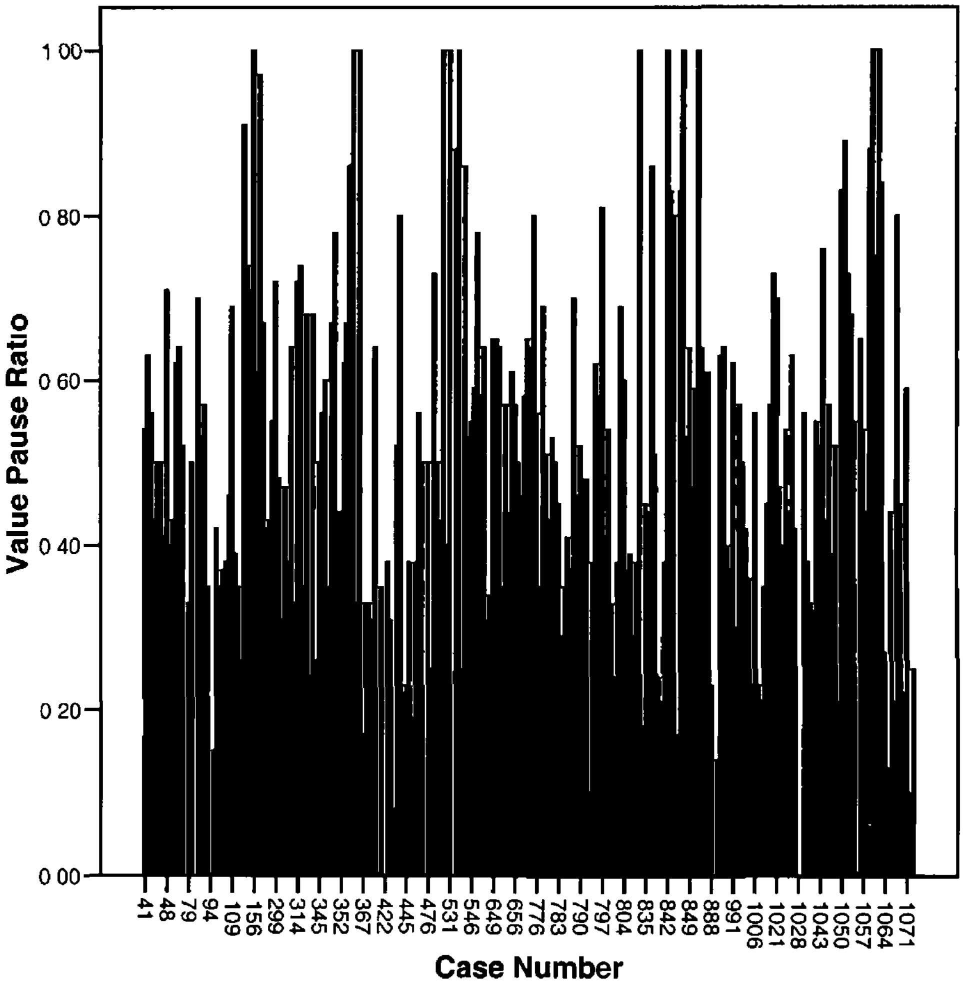


Figure 6 7 Pause Ratios for $S_{(min\ nti)}$

In both cases the data suggest that the majority of values lie below the 0 50 mark (i e 50% of the total processing time), with some values going as high as 1 00 (i e 100% of the processing time) A comparison of means and medians will allow us to make a meaningful comparison of the data for the two segment types

Segment Type	Mean Pause Ratio	Median Pause ratio
$S_{(nti)}$	5077	5000
$S_{(min\ nti)}$	4980	5000

Table 6 11 Median Pause Ratio Values

We can see that there is only a small difference between the mean values for the two segment types and that the median values are identical What this means is that the median pause ratio value for both segment types was 50%, i e for both segment types post-editors paused, on average, for 50% of the total processing time The standard deviation for $S_{(nti)}$ segments was 1980, and for $S_{(min\ nti)}$ segments it was higher at 2330

The data suggest that there are no significant differences for pause ratios between the two segment types (p= 0 548) This confirms the observation we made earlier in our discussion on pause behaviour where we stated that there was no discernible difference in the pause behaviour of post-editors for the two segment types Given this, along with findings from other researchers on the individual nature of pause behaviour (Hansen 2002, Alves 2006, Cenoz 2000), an analysis seeking to discover a correlation between pause ratios and specific NTIs would seem inappropriate On the subject of cognitive effort, we can conclude that our data show that there are no significant differences between the two segment types when pause ratio is used as a parameter

6.4.2 Choice Network Analysis

All data from the Choice Network Analysis are presented in Appendix H For each segment, the source text and NTIs are listed In the case of $S_{(min\ nti)}$ segment types, the classification "minimal NTIs" is used It has already been stated that the focus of this research was not to compare the human translated segments with the post-edited segments Nonetheless, the three human translated versions are presented with the CNA data in order to facilitate possible future research Having reference to the human translated versions also contributed to an understanding of what motivated certain changes implemented by the post-editors [70] Following this, and as indicated in Chapter 5, a table comparing all post-edited target segments with the MT output is given and the nodes where differences occur are highlighted in bold A commentary is then included for each segment which ends with a summary identifying what, if any, features in the source text appear to have caused post-editing effort In order to give an example of what our application of Choice Network Analysis looks like, we have included two examples here, one for a $S_{(min\ nti)}$ segment (segment number 38) and one for a $S_{(nti)}$ segment (segment number 46) For details on the CNA for each segment, the reader is referred to Appendix H [71]

6.4.2.1 SAMPLE CNA FOR $S_{(MIN-NTI)}$ SEGMENT

Source Text Sentence

In order to update the generated text, press the "lightning bolt" icon on the second toolbar

Negative Translatability Indicators

Minimal

Raw MT Output

[70] For example, in segment 15 two translators started the segment with "Weitere Informationen finden Sie " and this was also the case for six post editors The MT output was a literal translation "Sehen Sie "

[71] Note that for each CNA we also include a commentary on the changes that have been made, if any We omit the commentary here for reasons of space but they can be found under the relevant segment number in Appendix H

Um den generierten Text zu aktualisieren, drucken Sie das "Blitzbolzen" Symbol auf der zweiten Symbolleisten

Subject T1

Zum Aktualisieren des generierten Textes drucken Sie das Symbol "Blitz" in der zweiten Funktionsleiste

Subject T2

Wenn Sie den generierten Text andern mochten, klicken Sie das Symbol mit dem Blitz auf der zweiten Funktionsleiste an

Subject T3

Um den generierten Text zu aktualisieren konnen Sie auf das Symbol mit dem Blitz in der zweiten Funktionsleiste klicken

MT	Um den generierten Text zu aktualisieren,	drucken Sie das	**"Blitzbolzen" Symbol**	**auf der**	zweiten Symbolleisten
P1	Um den generierten Text zu aktualisieren,	klicken Sie auf das	**Blitzsymbol**	**in der**	zweiten Symbolleiste
P2	Um den generierten Text zu aktualisieren,	drucken Sie das	**Symbol mit dem Blitz**	**in der**	zweiten Symbolleiste
P3	Um den generierten Text zu aktualisieren,	drucken Sie das	**"Blitz-"Symbol**	**auf der**	zweiten Symbolleiste
P4	Um den generierten Text zu aktualisieren,	drucken Sie das	**Blitz-Symbol**	**auf der**	zweiten Symbolleiste
P5	Um den generierten Text zu aktualisieren,	drucken Sie das	**Blitzsymbol**	**in der**	zweiten Symbolleiste
P6	Um den generierten Text zu aktualisieren,	drucken Sie das	**Symbol "Blitz"**	**in der**	zweiten Symbolleiste
P7	Um den generierten Text zu aktualisieren,	klicken Sie auf das	**Blitzableitersymbol**	**in der**	zweiten Symbolleiste
P8	Um den generierten Text zu aktualisieren,	drucken Sie das	**Symbol "Blitzbolzen"**	**in der**	zweiten Symbolleiste
P9	Zur Aktualisierung des generierten Texts	klicken Sie auf	**das "Blitz"-Symbol**	**auf der**	zweiten Symbolleiste

Table 6 12 Sample CNA for $S_{(min\ nti)}$ segment

6.4.2.2 SAMPLE CNA FOR $S_{(NTI)}$ SEGMENT

Source Text Sentence

Select the C Drive

Negative Translatability Indicators

Short Segment

Raw MT Output

Wahlen Sie das C- Laufwerk aus

Subject T1
Wahlen Sie das C- Laufwerk aus

Subject T2
Wahlen Sie hierzu zunachst Laufwerk C,

Subject T3
Wahlen Sie das Laufwerk C aus

MT	Wahlen Sie	das	**C- Laufwerk**	aus
P1	Wahlen Sie	-	**Laufwerk C**	aus
P2	Wahlen Sie	das	**Laufwerk C**	aus
P3	Wahlen Sie	das	**C- Laufwerk**	aus
P4	Wahlen Sie	das	**C- Laufwerk**	aus
P5	Wahlen Sie	das	**C- Laufwerk**	aus
P6	Wahlen Sie	das	**Laufwerk "C "**	aus
P7	Wahlen Sie	das	**Laufwerk C**	aus
P8	Wahlen Sie	das	**Laufwerk C**	aus
P9	**Wechseln Sie zu**	-	**Laufwerk C**	-

Table 6 13 Sample CNA for $S_{(nti)}$ segment

6.4.3 Discussion of CNA Results

Here we will structure the discussion according to the following questions

- Which NTIs appear to impact on post-editing effort?
- Which NTIs appear *not* to impact on post-editing effort?
- When segments are classified as having "minimal NTIs", what features require post-editing?

Before proceeding, we will define what we mean by "impact on post-editing" an NTI is deemed to have had an impact if the MT output for that NTI has been changed by two or more post-editors In undertaking this analysis, it was deemed appropriate to introduce "degrees of impact" in order to get an overview of the extent to which specific NTIs were problematic Thus, we have introduced three categories of impact "high impact" is when 50% or more of the occurrences of a specific NTI have had an impact on post-editing, "moderate impact" is when between 31% and 49% of occurrences of a specific NTI have had an impact on post-editing, and "low impact" is when 30% or fewer occurrences have had an impact

6.4.3.1 NTIs with High Impact on Post-Editing Effort

Table 6 14 lists those NTIs that have had a high impact on post-editing effort

NTI	Total Number in ST	Number in segments with high impact on post-editing	% of this NTI occurring in segments with high impact on post-editing
Gerund	24	17	70%
Ungrammatical construct	3	2	67%
Post-modifying adjectival phrase	2	2	100%
Use of (s) for plural	3	2	67%
Non-finite verb	4	2	50%
Slang	2	2	100%
Misspelling	3	3	100%
Long Noun Phrase	10	7	70%
Ellipsis	9	7	78%
Long Sentence	3	2	67%
Verbs with Particles	2	2[72]	100%
Use of Footnotes	2	2	100%
Multiple Prepositions	2	1	50%
Short segment	6	4	67%

Table 6 14 NTIs with High Impact on Post-Editing

[72] The full ST contains two occurrences of verbs with particles The second occurrence is in segment 158 which, for reasons explained earlier, was not analysed A cursory analysis of the MT output for this segment suggests that the NTI would also impact on post editing effort here as the MT output is of a low quality (*Find and paste in artwork* is translated as *Finden Sie alle Suchen und Paste in Kunstwerk*) Therefore, we have included this second occurrence here too

6.4.3.2 NTIs with Moderate Impact on Post-Editing Effort

Table 6 15 lists those NTIs that have had a moderate impact on post-editing effort

NTI	Total Number in ST	Number in segments with moderate impact on post-editing	% of this NTI occurring in segments with moderate impact on post-editing
Multiple Coordinators	2	1	50%
Punctuation	15	8[73]	53%
Passive Voice	5	2[74]	40%
Not an full syntactic unit	5	2	40%
Personal Pronouns	16	7	44%
Use of slash as a separator	3	1[75]	33%
Ambiguous scope in coordination	3	1	33%
Parentheses	3	1	33%
Proper Noun	25	8	32%
Missing Relative Pronoun "that"	3	3	100%

Table 6 15 NTIs with Moderate Impact on Post-Editing

6.4.3.3 NTIs With Low Impact on post-editing effort

Table 6 16 lists those NTIs that have had a low impact on post-editing effort

NTI	Total Number in ST	Number in segments with low impact on post-editing	% of this NTI occurring in segments with low impact on post-editing
Abbreviation	18	1[76]	5%
Demonstrative Pronoun	7	2	29%
Missing "in order to"	5	0	0%
Contraction	2	0	0%

Table 6 16 NTIs with Low Impact on Post-editing

6.4.3.4 Non-NTIs and their impact on post-editing

As can be determined from the Choice Network Analysis for many segments, post-editing does not only occur around NTIs We have already drawn attention to this under the

[73] There are numerous different punctuation signs involved in the seven cases where post editing was not impacted e g missing period semi-colon, comma and colon Of the segments where post editing was impacted seven out of the eight contain a semi colon suggesting that this punctuation sign is particularly problematic

[74] However, in segment 31 some doubt remained as to whether post editing effort was caused by the passive voice (see the CNA for this segment) Also for segment 54 the changes made are relatively minor

[75] In the segment where post editing was impacted (segment 59) the slash separates two nouns In those segments where post editing was not impacted (14 and 24), the slash separates an acronym and number – OS/2 – and two coordinators – and/or This suggests that the use of the slash is most problematic when is separates two nouns, especially in inflected languages (see the CNA for segment 59)

[76] Even in the one instance where post editing occurs around the abbreviation (segment 26), this is most likely due to the "like" that is appended to the abbreviation (*WYSIWYG like*), rather than to the abbreviation itself

section on Relative Post-Editing Effort and will expand on it here using our observations from CNA It should be remembered here that in preparing the ST for machine translation, two CL checkers were used The features under discussion in this section were not highlighted as problems by either of these CL checkers

As already indicated, the detailed Choice Network Analysis for each segment is available in Appendix H Therefore, we will not reproduce or discuss every feature in detail here Rather, we would like to categorise and summarise the most obvious features and reference will be provided to individual segments as examples of these

Terminological Problems

Post-editing was necessary for a considerable number of computer-specific lexical items For example, *online delivery* (segment 4a), *double-click* (segment 9), *Enter key* (segment 100), *Save button* (segment 119) The reader is reminded that the original ST comprised 1,777 words The 79 terms coded represent just 4% of the total word count It would not be unreasonable to expect that post-editing effort would have been reduced if more terms had been coded

Verbs

Considerable post-editing effort occurred around verb forms For example, the data illustrate several changes made to verbs such as *have* (segment 25), *be* (segment 85), *appears* (segment 35), *includes* (segment 62), *opens* (segment 33), to name just a few In such cases, the MT-generated verb is not incorrect, i e the meaning is correct and some post-editors accepted the verb proposed by the MT system However, other post-editors instigated changes that could be motivated by stylistic preferences and/or the compulsion towards alternative TL collocations For example, in segment 25 the ST reads *Epic has an intagrated Table Editor* (one of the NTIs is a misspelling) This segment was translated by the MT engine as *Epic hat einen intagrated Tabelleneditor* Five post-editors accepted *hat*, but four felt compelled to change this to *verfugt uber* Likewise in segment 85, *ist* is changed to *handelt es sich um* by two post-editors It is probable that what we are witnessing here is years of professional training and experience at work where the translator strives for a stylistically acceptable translation and where certain TL collocations prevail over what might be seen as inferior, but accurate, translations

Formulaic Phrases

The phrase *See X for more information* was frequently the focus of post-editing effort (see, for example, segments 14 and 121) Other established phrases in English which also caused post-editing effort include *in the following ways* (segment 8) and *as large as/as small as* (segment 28)

Prepositions

Although “multiple prepositions” is mentioned above as an NTI, the occurrence of single prepositions is not usually deemed problematic in CL [77] Nonetheless, there are several examples of post-editing effort around single prepositions in the data presented here - for example, *after* (segment 32), *of* (segment 94), *as* (segment 95), *in* (segment 115) and *from* (segment 177)

Polysemy

Polysemy is a well-known problem for machine translation and this was no exception during this research In particular, the polysemous nature of the noun *letter* (segments 28 and 61) and the verb *order* (segment 35) caused problems

Spelling Rules

The new spelling rules for German were clearly not programmed into the MT system and this resulted in considerable post-editing effort whereby some post-editors, who subscribed to the new spelling rules, changed lexical items such as *daß* and *muß* to *dass* and *muss* It is interesting to note that not all post-editors made these changes

6.4.4 Conclusions on Cognitive Effort

One of our findings within the parameter of cognitive effort is that there was no significant difference between the ratio of pause to keyboarding for the two segment types

A second finding has to do with methodology we found that an analysis of the position and duration of pauses within post-editing activity did not contribute to our understanding of the relationship between cognitive effort and NTIs

Using Choice Network Analysis as the method for assessing cognitive effort, our study has shown that some NTIs impact on post-editing effort more than others Those NTIs have been listed above under section 6 4 2 according to the criteria of high, moderate and low impact These data could be used as a guide for selecting CL rules that are highly effective in reducing post-editing effort, at least for the English-German language pair in the IT domain

We have also identified additional linguistic features that impacted on cognitive effort These include computer-specific terms that were not in the MT dictionary, general verbs,

[77] Rules from the CL rule sets analysed for this study generally refer to the need to avoid the use of multiple prepositions, or to specify what prepositions may be used and in the case of Attempto how they may be used

some formulaic expressions in English, prepositions, polysemy and German spelling rules Post-editing effort could conceivably be reduced if the specialised terms, formulaic expressions and polysemous words were coded correctly in an MT dictionary In addition, the new German spelling rules could be programmed in an MT engine and this, in turn, would reduce post-editing effort The post-editing effort expended on changing verbs and prepositions is perhaps more difficult to deal with as many of these changes appear to have been motivated by stylistic and/or TL collocational preferences Again, some dictionary coding where favoured translations of verbs could be given precedence in an MT dictionary might reduce post-editing effort Training post-editors not to make changes when translations of verbs are accurate and acceptable would also address this issue

6.5 Data Analysis Conclusions

From a purely temporal point of view, post-editing was faster than translation in this study The processing speeds for segments with minimal negative translatability indicators were significantly faster than for segments containing one or more negative translatability indicators The measure of Relative Post-Editing Effort confirms that post-editing required, on average, less effort than translation Although we found no statistically significant differences in the RPE values of the two segment types, a higher proportion of $S_{(\mathrm{min\ nti})}$ segments (42%) had lower RPE values than $S_{(\mathrm{nti})}$ segment types (20%) We should add, however, that some $S_{(\mathrm{min\ nti})}$ segments were clearly not free from post-editing effort

On technical effort, the data show significant differences in the number of deletions and insertions required for $S_{(\mathrm{min\ nti})}$ and $S_{(\mathrm{nti})}$ segment types No such differences have been shown to exist for the action of cutting and pasting, but because the post-editors did not use this technique very often, data are too sparse to draw firm conclusions anyway An appropriate post-editing environment that would make cutting and pasting easier might reduce post-editing effort overall In addition, training on how to effectively recycle the words and phrases produced by the MT system would help to reduce post-editing effort

The data imply that there are no significant differences in cognitive effort between the two segment types when pause ratio is used as the parameter of measurement An analysis of pause position and location did not lead to any fruitful conclusions on cognitive effort The results from Choice Network Analysis have demonstrated that some NTIs seem to require more cognitive effort than others (see 6 4 3 1 and 6 4 3 2 for the list) There is an interesting correlation between the list of NTIs that appear in segments with a low processing speed, a high RPE value and the NTIs that are listed has having either a “moderate” or “high” impact on cognitive effort (defined in terms of the percentage of occurrences that led to post-edits) Table 6 17 shows these correlations

NTI	Low Processing Speed	High RPE Value	High Impact on Post-editing	Moderate Impact on Post-editing
Gerund	Yes	Yes	Yes	-
Proper Noun	Yes	Yes	-	Yes
Problematic Punctuation	Yes	Yes	-	Yes
Ungrammatical Construct	Yes	Yes	Yes	-
Use of (s) for plural	Yes	Yes	Yes	-
Non-Finite Verb	Yes	Yes	Yes	-
Not a full syntactic unit	Yes	Yes	-	Yes
Long NP	Yes	Yes	Yes	-
Short Segment	Yes	Yes	Yes	-

Table 6 17 Correlations between NTIs and Post-editing Effort

All NTIs in the high RPE list appear in either the moderate impact or high impact lists Furthermore, several of the linguistic features identified as being problematic for post-editing (e g terminology, verbs, formulaic language etc), also occur in the high RPE value list There also appears to be some correlation with the low RPE value list, although this is less striking all NTIs in the low RPE value list appear in either the moderate impact or low impact on post-editing list, with one exception (ellipsis) We can observe from this that the method of calculating Relative Post-editing Effort has triangulated well with Choice Network Analysis

In conclusion, we have seen that post-editing is faster than translation as a first-pass exercise, but we should once again mention that no quality analysis has been carried out on the resulting product Some measures of temporal effort (e g processing speed) and technical effort (e g number of insertions and deletions) suggest that $S_{(min\ nti)}$ segments are easier to post-edit However, other temporal measures (e g Relative Post-Editing Effort) and cognitive measures (e g pause ratio and CNA) suggest that there are no significant differences between the post-editing effort for the two segment types Therefore, we cannot conclusively say that the elimination of the NTIs under study will always lead to easier and faster post-editing However, we have identified those NTIs that have required the greatest post-editing effort and those that appear not to result in post-editing effort We have also identified other linguistic features that have added to post-editing effort (e g uncoded terminology, verbs, prepositions, polysemy, spelling rules), but which were not included in our NTI list and which, furthermore, were not identified as problems by either of the CL checkers used in the preparation stages More extensive terminology coding would have reduced post-editing effort However, stylistic and TL collocational preferences also seem to have played an important role in the post-editing process The identification of NTIs with high and low impact on post-editing effort as well as non-NTIs that have impacted on post-editing effort will hopefully contribute towards the fine-tuning of CL rules and the ultimate reduction of post-editing effort

Chapter 7

7. CONCLUSION

7.1 THE OBJECTIVES

The objective of this research was to investigate the hypothesis that by controlling the input text to MT, post-editing effort is reduced. A second, related objective was to see if we could identify correlations between specific CL rules and post-editing effort. As stated in our introduction, the assumption that controlling the input text reduces post-editing effort is a long-standing one in the domain of translation automation, but it has been largely uncontested – at least empirically. It is our contention that a subjective evaluation of MT output does not produce an accurate measurement of potential post-editing effort. For example, if an evaluator rates a sentence that has been machine translated as "excellent", does this mean that the sentence will not require post-editing? Here we agree fully with Ryan's contention (1988) that the effectiveness of MT must be measured not by the speed of the system but by the effectiveness of the post-editing process. We also remind ourselves here of a finding from Krings's (2001) research that MT output rated as "medium" quality requires more post-editing effort than that rated as "bad".

In order to meet the objective, Negative Translatability Indicators – NTIs – were identified and introduced into our source text. Once post-editing had been completed, we investigated the correlations between temporal post-editing effort and sentences with NTIs ($S_{(nti)}$) and those without ($S_{(min\ nti)}$). We then investigated the correlations between specific NTIs and temporal, technical and cognitive post-editing effort.

7.2 FINDINGS

Our data show that, when taken as a first-pass exercise with no revision or proof-reading and with the specified goal of producing publishable quality output, post-editing is, on average, faster than human translation. In addition, we have shown that segments with minimal occurrence of NTIs can be processed faster by post-editors than those containing NTIs. This finding provides empirical evidence to support the assertion that controlling the input to MT leads to faster post-editing. However, the findings are based on median values and it is clear that individuals differ in their average post-editing speeds.

By combining results from temporal, technical and cognitive measures of effort, we have also demonstrated that some NTIs, e.g. use of the gerund, ungrammatical constructs, use of "(s)" to mark the plural, non-finite verbs, long noun phrases and short segments, lead to higher post-editing effort than others. NTIs that lead to moderate post-editing effort include: proper nouns, problematic punctuation, and incomplete syntactic units. The data also demonstrate that some NTIs, e g abbreviations, demonstrative pronouns, missing "in order to", and use of contractions, do not have a high impact on post-editing effort These results demonstrate that CL rules do not have equal impact in reducing post-editing effort

In addition to meeting the two objectives outlined above, we also made some additional discoveries Choice Network Analysis helped us to identify linguistic features that were not captured in our list of NTIs where post-editing effort was required, e g terminology that was not coded in the MT dictionary, verb forms such as *have* or *includes*, and certain formulaic phrases, such as *See X for more information*, to name but some Our finding is that, even if we control the input to MT, post-editors make changes to sentences where we might not expect them to Post-editing effort could be reduced further by, for example, reducing the number of verbs that can be used in the ST, coding more terms in the MT dictionary, and even by including common formulaic phrases and their preferred translations in the MT dictionary On the topic of formulaic phrases, we observed that the standard way of saying things in certain domains (e g *Weitere Informationen uber X finden Sie in*) are so engrained in post-editors' minds that the urge to change reasonably intelligible MT output is too great to resist One could argue that the solution here is to train post-editors to fight this urge However, it might be more reasonable to suggest that MT systems should be developed further so that they can produce the correct translations for such formulaic phrases In that way, the post-editor would not have to even think about the possibility of changing the output and cognitive effort would be spared

Our analysis of dictionary usage showed differences in behaviour between the students who participated in the pilot project and the professional translators (post-editors) The post-editors barely used the dictionary look-up facility, while some of the student translators were over-reliant on it, confirming findings by Livbjerg and Mees (2003) that semi-professional translators can be over-reliant on dictionaries However, it is not clear as to whether the lack of use of the dictionary look-up facility by post-editors resulted from a lack of need on their behalf or from a lack of proficiency in use of the underlying technology (Translog)

Considerable thought was given to the effectiveness of pauses as indicators of cognitive effort Our finding on this subject was in keeping with what other translation process researchers have found (Jakobsen 2005, Alves 2006) while pauses can be seen as delimiters of cognitive rhythm, an analysis of pause duration and location does not produce reliable data on cognitive effort because pauses are subject to individual, erratic behaviour and also because we cannot reliably record what is going on in an individual's mind during a pause

The analysis of technical effort revealed that subjects prefer to re-type words that already exist in the target text than to cut and paste those words or recycle them into different inflected forms of the same word Perhaps subjects made the assumption that there is less effort involved in re-typing a word or phrase in one part of a sentence than cutting that word or phrase from another part of the sentence and pasting it elsewhere? Alternatively, post-editors might be so pre-occupied with generating a revised target text that they cannot give attention to cutting and pasting? This topic would be worthy of further investigation

7.3 COMMENTS ON METHODOLOGY

To our knowledge, this is the first time that Translog has been used in triangulation with CNA in an empirical study of post-editing effort By pioneering this approach, we hope to have made a contribution regarding the methodologies that can be employed for research into post-editing and, by extension, translation processes It is our hope that these two methodologies will be adopted by other researchers in the future in order to make further contributions to the field

The triangulation of Translog with Choice Network Analysis provided us with a good methodological framework with which to investigate our hypotheses A small number of minor weaknesses in Translog were identified With the possible exception of the dictionary look-up functionality, which may have had an impact on post-editors' dictionary look-up behaviour as discussed above, we feel that the weaknesses were minor and did not impact on the study overall

With regard to Choice Network Analysis, some reservations still exist about the theory underpinning this methodology In Hale and Campbell (2002 15), the rationale for counting the number of renditions produced by translators (or, in our case, post-editors) "is that the different renditions represent the options available to a group of subjects, and that each subject is faced with making a selection from those options" Can we accept that if one post-editor produces an alternative rendition of a TT segment, that alternative was available to all post-editors and that the two were actively considered by all subjects? What about alternatives that were considered, but rejected, which therefore do not appear in the TT? CNA has no way of accounting for those unmanifested alternatives In the same way that think-aloud protocols cannot provide us with access to all the thoughts that might go through a subject's mind while they are working (because of automatisation and individual propensity to verbalise), CNA cannot provide us with evidence of all the alternatives an individual may have considered while producing a TT, nor can it provide us with evidence that all subjects considered all alternatives On these questions one might counter, however, that it is precisely the careful selection of a homogenous group of subjects (who share similar linguistic, educational and professional backgrounds), that allows us to assume a certain homogeneity in subjects' (text-) linguistic competence and thus their potential responses to the MT output with which they are faced Because of the homogeneity of the group, we have good reason to assume that choices manifest in the output of individual editors were also available to other editors, even if they did not select them On the issue of options that are not made manifest at all, i e that do not appear in the output of any of the editors, we can say that the non-selection of particular wordings does not rule these out as potential wordings, (i e absences are not interpreted as indicating impossibilities), but that our methodology, like other empirical methodologies in, e g linguistics, by its very nature focuses on actual, attested, evidence

The application of CNA raises another issue, highlighted by Breedveld (2002) who argues that cognitive activities in translation differ according to the moment and context in which they occur

> *as the text-produced-so-far evolves, the writer must continually re-represent the task as well as the text*
>
> *(ibid 225)*

In the context of CNA and post-editing, the question therefore emerges to what extent do choices made earlier in the post-editing of a sentence influence choices made later in the post-editing of the same sentence? In our experience, CNA causes the researcher to focus on chunks of text and we make assertions about text difficulty based on the number of alternatives produced However, the initial choices made by a post-editor could lead to necessary changes later in a sentence Alternative renditions could, therefore, be motivated by previous choices rather than by text difficulty This suggests that CNA ought to take into account not just the number of alternative renditions, but also the domino effect that the selection of any one rendition might have

The sample diagrams on Choice Network Analysis provided by Campbell (1999, 2000a, 2000b) and Campbell and Hale (1999) were difficult to replicate and we felt that their replication for the number of segments and post-editors in our study would be too cumbersome and time-consuming Therefore, we opted for a tabular layout for CNA and this proved to be effective

Despite the weaknesses highlighted above, the fact remains that we found CNA to be a useful methodology in this study – it provided us with a systematic framework for recording the alternatives chosen by post-editors and pointed to those parts of the TT that the post-editors thought needed repairing It allowed us to establish correlations between post-editing effort – technical, temporal, and cognitive – and NTIs

In Chapter 6, we drew attention to the fact that two subjects opted out of the experiment at different points, by producing unintelligible output Neither subject gave any indication, either before or after the experiment, that they were unhappy with participating Their change in behaviour was only discovered after considerable effort had been put into analysing their data While the use of thinking-aloud might have discouraged this "opting out", or might, at least, have identified the fact that it happened earlier in the data analysis cycle, we ruled it out because of the evidence provided by researchers that it would interfere with the process and slow it down The alternative of retrospective thinking-aloud was also ruled out for two reasons (1) this decision was founded on the claim (Bernardini 2001 243, following Ericsson and Simon) that retrospective protocols can often be incomplete because the thought processes involved in the task being analysed have moved into long-term memory by the time a retrospective protocol is recorded and (2) it was anticipated that the time required for analysis of data using Translog and Choice Network Analysis would be significant in itself and, therefore, the transcription of a retrospective protocol, along with the classification and analysis of data that this would involve, was considered to be excessively time-consuming However, a retrospective protocol would at least have identified the issue of a change in behaviour by two subjects at an earlier stage in the data analysis cycle and may also have provided an explanation for this change in behaviour For future similar research efforts, we would recommend at least a retrospective interview of subjects – a sort of "debriefing" session - if not an entire retrospective protocol

7.4 FUTURE RESEARCH

As mentioned in Chapter 3, a recent report on best practices in post-editing (Joscelyne 2006) reported average daily throughput rates for publishable quality post-editing as being approximately 5,250 words per day The average post-editing throughput in this study was 7,036 words per day [78] However, an important question must be asked here can post-editors really work for eight hours per day? The task of post-editing is reported, albeit anecdotally, as being tiresome We can speculate that a human would probably not be able to post-edit for eight hours in one day Therefore, the throughput rate as reported by Joscelyne (2006) of approximately 5,000 words per day is probably more realistic [79] Further research into cognitive effort and post-editing is important in order to avoid unrealistic expectations on throughput rates in the industry The use of eye-tracking methodologies (Duchowski 2003) to

[78] Here, the rate of words per minute recorded in this study have been converted to words per hour and an assumption has been made that there are eight working hours in a day

[79] As a matter of interest Joscelyne also reports an average throughput of 10 500 words per day for "light" post editing

investigate translation processes and cognitive effort and translation technology, in particular, has already begun [80]

Due to limitations on time and the capacity of one researcher, the findings reported here are limited to one MT system, language pair and direction, and text type The number of subjects involved and the number of occurrences of NTIs was also necessarily limited Future research could involve the extension of this study to other MT engines, language pairs and directions and text types The number of occurrences of NTIs could be increased as could the number of subjects (time-permitting, of course)

Throughout this study we have drawn attention to the fact that we have not investigated the relationship between post-editing effort, NTIs and the quality of the post-edited product Nor have we investigated the quality of the post-edited product compared with that of the translated product Given that we have confirmed that post-editing was faster than translation, the logical follow-up question is what deficit, if any, exists between the post-edited and translated products?

In our section on Findings above, we drew attention to the observed differences in dictionary usage between student translators and our professional subjects Whether this resulted from differences in strategies or from a lack of trust in the Translog dictionary environment on the part of the professional subjects is something that would be worthy of further investigation

Finally, the observed unwillingness on the part of post-editors to cut and paste or recycle parts of words is a topic that may be worthy of future research Can this behaviour be replicated with another group of subjects? If so, does it result from a lack of post-editing experience (and can it therefore be changed through training) or does it result from a limitation in cognitive processing capabilities? Linked to that is the topic of computer-aided post-editing (CAPE), about which we are convinced we will hear more in the future, especially if the drive for translation automation continues How could a CAPE environment support the recycling of text? What are the requirements for such an environment? And, considering that post-editing of MT output is increasingly likely to occur in TM environments, what impact will this have on the functionality of TM tools? These are some of our suggestions for future research and we are hopeful that some (if not all) of the topics above will be the focus of research into the future

[80] The researcher has already embarked on some preliminary studies of cognitive effort (measured by pupil dilations) and translation in the SDLX/Trados Translator s Workbench environment using the Tobii eye tracker See also the EU funded project "Eye2IT" at http //http //www dpmi tu graz ac at/eye2it html (last accessed 23/04/2006)

We believe that this study has demonstrated that while the application of CL rules can reduce post-editing effort, the relationship is not a simple one Not all CL rules will have an equal effect on post-editing effort Also, even though NTIs have been removed from source sentences, some post-editing effort may still be required The implementation of a CL/MT solution appears to warrant a long-term study of the effectiveness of specific CL rules and of the linguistic features that are causing post-editing effort, but which are not controlled by CL rules

REFERENCES

Adriaens, Geert (1996), SECC: Using Text Structure Information to Improve Checker Quality and Coverage, in Adriaens et al. (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp. 226-232.

Adriaens, Geert (1994), Simplified English Grammar and Style Correction in an MT Framework: the LRE SECC Project, in *Proceedings of Translating and the Computer 16*, Aslib: London, UK, pp 78-88.

Adriaens, Geert, Jeffrey Allen, Arendse Bernth, Kurt Godden, Teruko Mitamura, Eric Nyberg, Rick Wojcik, Remi Zajac (2000) (eds), *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000),* Seattle, Washington.

Adriaens, Geert, R. Havenith, R. Wojcik, B. Tersago (1996) (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium.

Adriaens, Geert and L. Macken (1995), Technological Evaluation of a Controlled Language Application: Precision, Recall, and Convergence Tests for SECC, in *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI95),* Centre for Computational Linguistics, Leuven, Vol. 1, pp 123-141.

Allen, Jeffrey (2003), Post-Editing, in Somers, H. (ed), *Computers and Translation: A Translator's Guide*, Amsterdam/Philadelphia: John Benjamins, pp. 297-317.

Allen, Jeffrey (2002), Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes, Book review in *Multilingual Computing and Technology*, Volume 13, Issue 2, March, pp. 27-29.

Allen, Jeffrey and C. Hogan (2000), Toward the Development of a Postediting Module for Raw Machine Translation Output: A Controlled Language Perspective, in Adriaens et al. (eds) *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000),* Seattle, Washington, pp 62-71.

Almquivst, Ingrid and Anna Sågvall Hein (1996), Defining Scania Swedish - a Controlled Language for Truck Maintenance, in Adriaens et al. (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp.159-165.

Alves, Fabio (2006), A Relevance-Theoretic Approach to Effort and Effect in Translation: Discussing the Cognitive Interface between Inferential Processing, Problem-Solving and Decision-Making, in *Proceedings of the International Symposium on New Horizons in Theoretical Translation*

Studies, 19-20 January, Department of Translation, The Chinese University of Hong-Kong, pp 1-12

Alves, Fabio (ed) (2003), *Triangulating Translation Perspectives in Process Oriented Research*, Amsterdam John Benjamins

Alves, Fabio and José Luiz Gonçalves (2003), A Relevance Theory Approach to the Investigation of Inferential Processes in Translation, in Alves, Fabio (ed), *Triangulating Translation Perspectives in Process Oriented Research*, Amsterdam John Benjamins, pp 3-24

Arai, Yoshinori, Atsuo Shimada, Masumi Narita, User Behaviour During Post-Editing in Machine Translation, *IPSG SIGnotes Natural Language*, No 0 84

Arnold, Doug, L Balkan, L Humphreys, S Meijer, L Sadler (1994), *Machine Translation An Introductory Guide*, Oxford NCC Blackwell

AeroSpace and Defence Industries Association of Europe (ASD), ASD Simplified Technical English, ASD Specification ASD-STE 100™, [http //www simplifiedenglish-aecma org/Simplified_English htm], [Last accessed 10/04/2006]

Atwater, Kathleen (1998), Nortel Standard English as a Quality and Reliability Tool Distributed Report Ottawa, Canada Public Carrier Networks Information Development, Nortel Paper presented at the 1998 IEEE conference (no page numbers)

Baker, Kathryn, A Franz, P Jordan, T Mitamura, E Nyberg (1994), Coping with Ambiguity in a Large-Scale Machine Translation System, in Nagao, M (ed), Proceedings of the 15th International Conference on Computational Linguistics, Vol 1, Kyoto, Japan, August 5th-9th, pp 90-94

Barthe, Kathy (1998), GIFAS Rationalised French Designing One Controlled Language to Match Another, in Mitamura et al (eds) *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 87-102

Bédard, Claude (1992), La Prétraduction Automatique, Outil de Productivité et d'Evolution Professionnelle, in *Meta*, Vol 4, pp 738-760

Bennett, Scott (2000), Taking the Babel Out of Babelfish, *Language International*, Volume 12, no 3, Amsterdam John Benjamins, pp 20-21

Bernardini, Silvia (2001), Think-Aloud Protocols in Translation Research - Achievements, Limits, Future Prospects, in *Target*, 13 2, pp 241-263

Bernth, Arendse (2000), EasyEnglish Grammar Checking for Non-Native Speakers, in Adriaens et al (eds), *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*, Seattle, Washington, pp 33-42

Bernth, Arendse (1999a), Controlling Input and Output of MT for Greater User Acceptance, in *Proceedings of the 21st Conference on Translating and the Computer*, ASLIB, London, (no page numbers)

Bernth, Arendse (1999b), EasyEnglish A Confidence Index for MT, in *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation*, Chester College, England, pp 120-127

Bernth, Arendse (1998a), EasyEnglish Addressing Structural Ambiguity, in *Proceedings of the Third Conference of the Association of Machine Translation in the Americas- AMTA-1998*, Langhorne, PA, Association for Machine Translation in the Americas, pp 164-173

Bernth, Arendse (1998b), EasyEnglish Preprocessing for MT, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 30-41

Bernth, Arendse (1997), EasyEnglish A Tool for Improving Document Quality, in *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Washington, DC, pp 159-165

Bernth, Arendse and Claudia Gdaniec (2001), MTranslatability, *Machine Translation*, Vol 16, No 3, pp 175-218

Bernth, Arendse, and Michael McCord (2000), The Effect of Source Analysis on Translation Confidence, in John White (ed), *Envisioning Machine Translation in the Information Future, Proceedings of the 4th Conference of the Association for Machine Translation in the Americas- AMTA-2000*, 10th-14th October, Cuernavaca, Mexico, Springer, pp 89-99

Breedveld, Hella (2002), Translation Processes In Time, in *Target*, 14 2, pp 221-240

Byrne, Jody (2004), *Textual Cognetics and the Role of Iconic Linkage in Software User Guides*, PhD thesis, Dublin City University

Campbell, Stuart (2000a), Choice Network Analysis in Translation Research, in Olohan, Maeve (ed), *Intercultural Faultlines Textual and Cognitive Aspects – Research Models in Translation Studies I*, Manchester St Jerome, pp 29-42

Campbell, Stuart (2000b), Critical Structures in the Evaluation of Translations from Arabic into English as a Second Language, in *The Translator*, Volume 6, Number 2, pp 211-229

Campbell, Stuart (1999), A Cognitive Approach to Source Text Difficulty in Translation, in *Target*, 11:1, pp 33-63.

Campbell, Stuart and Sandra Hale (1999), What Makes a Text Difficult to Translate?, in *Proceedings of the 1998 ALAA Congress*, 30th June-3rd July, Griffith University, Queensland, Australia, available at http //www cltr uq edu au/alaa/proceed/camphale html [Last accessed 20/04/2006]

Cascales Ruiz, Remedios and Richard Sutcliffe (2003), A Specification and Validating Parser for Simplified Technical Spanish, in *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (CLAW 2003)*, 15th-17th May, Dublin City University, Ireland, pp 35-44

Cenoz, Jasone (2000), Pauses and Hesitation Phenomena in Second Language Production, ITL-Review of Applied Linguistics, vol 127-128, pp 53-69

Cohen, Louis and Michael Holliday (1982), Statistics for Social Scientists An Introductory Text with Computer Programs in BASIC, London and New York Harper and Row

Cremers, Lou (2001), Towards an Automated Translation Workflow at Oce Technologies, in *International Journal for Language and Documentation*, Issue 9, May/June, pp 24-25

Danks, Joseph, Gregory Shreve, Stephen Fountain, and Michael McBeath (eds) (1997), *Cognitive Processes in Translation and Interpreting*, California and London Sage Publications

Dechert, Hans W and Ursula Sandrock (1986), Thinking-Aloud Protocols, the Decomposition of Language Processing, in Cook, V (ed), *Experimental Approaches to Second Language Learning*, Oxford-New York Pergamon Press, pp 111-126

Douglas, Shona and Matthew Hurst (1996), Controlled Language Support for Perkins Approved Clear English (PACE), in Adriaens et al (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96)*, Centre for Computational Linguistics, Leuven, Belgium, pp 93-105

Duchowski, Andrew (2003), *Eye Tracking Methodology – Theory and Practice*, London Springer-Verlag London

EAMT/CLAW (2003), Controlled Language Translation, *Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop*, 15th-17th May, Dublin City University, Ireland

Ericsson, K A and H A Simon (1984), *Protocol Analysis Verbal Reports as Data*, Cambridge, Massachusetts MIT Press

Falkenheimer, E (1992) *Die Rolle der Praedition im Bereich des maschinellen Ubersetzens – Bestandsaufnahme und empirische Untersuchung*, Unpublished *Diplom* thesis, Hildesheim University of Hildesheim, Institut fur Angewandte Sprachwissenschaft

Farrington, Gordon (1996), AECMA Simplified English an Overview of the International Aircraft Maintenance Language, in Adriaens et al (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp 1-21

Foulin, Jean-Noel (1995), Pauses et Débits Les Indicateurs Temporels de la Production Ecrite, *L'Année Psychologique*, 95, pp 483-504

Frey, Lawrence R , Carl H Botan, Gary L Kreps (1991), *Investigating Communication - An Introduction to Research Methods*, Englewood Cliffs, N J Prentice Hall

Fuchs, Norbert, Uta Schwertel, Rolf Schwitter (1999), Attempto Controlled English Language Manual, Version 3 0, Institut fur Informatik der Universitat Zurich, Nr 99 03

Fuchs, Norbert and Rolf Schwitter (1996), Attempto Controlled English (ACE), in Adriaens et al (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp 124-136

Fuchs, Norbert, and Rolf Schwitter (1995), Attempto Controlled Natural Language for Requirements Specifications, in *Proceedings of the7th IIPS 95 Workshop on Logic Programming Environments*, Portland Oregon, (no page numbers)

Gdaniec, Claudia (1994), The Logos Translatability Index, in *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, 5th-8th October, Columbia, Maryland, USA, pp 97-105

Gennrich, K (1992), *Die Nachredaktion von Maschinenubersetzungen am Bildschirm - eine Prozeßuntersuchung*, Unpublished *Diplom* thesis, Hildesheim University of Hildesheim, Institut fur Angewandte Sprachwissenschaft

Gerloff, Pamela (1988), From French to English A Look at the Translation Process in Students, Bi-Linguals and Professional Translators, Unpublished dissertation, Cambridge, MA, Harvard University

Gerloff, Pamela (1987), Identifying the Unit of Analysis in Translation Some Uses of Think-Aloud Protocol Data, in C Faerch and G Kasper (eds), *Introspection in Second Language Research*, Clevedon Multilingual Matters, pp 135-158

Gilhooley, K J (1987), Individual Differences in Thinking-Aloud Performance, in *Current Psychological Research and Reviews*, 5, pp 328-334

Glover, Alan, Edward Johnson, Fred Weeks, Peter Strevens (1984), *Seaspeak Reference Manual*, Pergamon Press

Godden, Kurt (2000), The Evolution of CASL Controlled Authoring at General Motors, in Adriaens et al (eds), *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*, Seattle, Washington, pp 14-19

Godden, Kurt (1998), Controlling the Business Environment for Controlled Language, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 185-189

Goukens, Loude (1999), Languaging the Web Loude Goukens talks to Lernout and Hauspie's Marc Bautil about how iTranslator will get its translation engines to pull together, *LE Journal The Journal of Record for Human Language Technology*, July [no page numbers], available at http //www crux be/ArticleLH PDF [Last accessed 30/01/2006]

Green, Roy (1982), The MT Errors Which Cause Most Trouble to Post-Editors, in Veronica Lawson (ed), Practical Experience of Machine Translation, Proceedings of a Conference, London, 5-6 November, pp 101-104

Hale, Sandra and Stuart Campbell (2002), The Interaction Between Text Difficulty and Translation Accuracy, in *Babel*, 48 1, pp 14-33

Haller, Johann (2000), MULTIDOC - Authoring Aids for Multilingual Technical Documentation, *First Congress of Specialized Translation*, Barcelona, March, (no page numbers)

Hansen, Gyde (2003), Controlling the Process Theoretical and Methodological Reflections on Research into Translation Processes, in Fabio Alves (ed), *Triangulating Translation Perspectives in Process Oriented Research*, Amsterdam John Benjamins, pp 25-42

Hansen, Gyde (ed) (2002), *Empirical Translation Studies - Process and Product*, Copenhagen Studies in Language 27, Copenhagen Samfundslitteratur

Hansen, Gyde (ed) (1999a), Probing the Process in Translation Methods and Results, Copenhagen Studies in Language 24, Copenhagen Samfundslitteratur

Hansen, Gyde (1999b), Das kritische Bewußtsein beim Ubersetzen Eine Analyse des Ubersetzungsprozesses mit Hilfe von Translog und Retrospektion, in Hansen, G (ed) *Probing the Process in Translation Methods and Results*, Copenhagen Studies in Language 24, Copenhagen Samfundslitteratur, pp 43-68

Somers, Harold (ed) (2003), *Computers and Translation A Translator's Guide*, Amsterdam John Benjamins

Harris, Zellig (1968), *Mathematical Structures of Language*, Huntington, New York Krieger Publishing Company

Hirschmann, L. and N. Sager (1982), Automatic Information Formatting of a Medical Sublanguage, in Kittredge, R. and J. Lehrberger (eds) *Sublanguage: Studies of Language in Restricted Semantic Domains*, Berlin, New York: de Gruyter, pp. 27-80.

Hoard, James, Rick Wojcik, Katherina Holzhauser (1992), An Automated Grammar and Style Checker for Writers of Simplified English, in Patrik Holt and Nole Williams (eds), *Computers and Writing: State of the Art*, Dordrecht: Kluwer Academic Publishers, pp. 278-296.

House, Juliane (1988), Talking to Oneself or Thinking with Others? On Using Different Thinking-Aloud Methods in Translation, in *Fremdsprachen Lehren und Lernen*, 17, pp. 84-98.

Huijsen, Willem-Olaf (1998), Controlled Language - An Introduction, in Mitamura et al. (eds), *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp. 1-15.

Hutchins, W. John and Harold L. Somers (1992), *An Introduction to Machine Translation*, London: Academic Press.

Jääskeläinen, Riitta (2002), Think-Aloud Protocol Studies into Translation – An Annotated Bibliography, in *Target*, 14:1, pp. 107-136.

Jääskeläinen, Riitta (1990), *Features of Successful Translation Processes: A Think-Aloud Protocol Study*, A licientiate thesis of the University of Joensuu, Savonlinna School of Translation Studies.

Jääskeläinen, Riitta (1987), What Happens in a Translation Process - Think-Aloud Protocols of Translation, Unpublished pro gradu study, Savonlinna: University of Joensu, Savonlinna School of Translation Studies.

Jääskeläinen, Riitta and Sonja Tirkkonen-Condit (1991), Automatised Processes in Professional vs. Non-Professional Translation: A Think-Aloud Protocol Study, in Tirkkonen-Condit, S. (ed), *Empirical Research in Translation and Intercultural Studies; Selected Papers of the TRANSIF Seminar, Savonlinna 1988*, Tübingen: Gunter Narr Verlag. pp. 89-109

Jakobsen, Arnt Lykke (2005), Instances of Peak Performance in Translation, in *Lebende Sprachen*, No. 3, pp. 111-116.

Jakobsen, Arnt Lykke (2003), Effects of Think Aloud on Translation Speed, Revision, and Segmentation, in Alves, Fabio (ed), *Triangulating Translation: Perspectives in Process Oriented Research*, Amsterdam: John Benjamins, pp. 69-95.

Jakobsen, Arnt Lykke (1999), Logging Target Text Production with Translog, in Hansen, Gyde (ed), *Probing the Process in Translation: Methods and Results*, Copenhagen Studies in Language 24, Copenhagen: Samfundslitteratur, pp. 9-20.

Jakobsen, Arnt Lykke (1998), Logging Time Delay in Translation, in *LSP Texts and the Translation Process, Copenhagen Working Papers*, pp 73-101

Janssen, Gerd, Gerhard Mark and Bernd Dobbert (1996), Simplified German - A Practical Approach to Documentation and Translation, in Adriaens et al (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp 150-158

Johnson, Edward (1996), LinguaNet™ - Controlling Police Communication, in Adriaens et al (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp 115-123

Johnson, Roderick L and Peter Whitelock (1987), Machine Translation as an Expert Task, in Nirenburg, S (ed), *Machine Translation Theoretical and Methodological Issues,* New York, Cambridge Cambridge University Press, pp 136-144

Joscelyne, Andrew (2006), *Best-Practices in Post-Editing,* Translation Automation User Society (TAUS) Special Report (available to TAUS members only), http //http //www translationautomation com/ [Last accessed 19/04/2006]

Kamprath, Christine, Eric Adolphson, Teruko Mitamura, Eric Nyberg (1998), Controlled Language for Multilingual Document Production Experience with Caterpillar Technical English, in Mitamura et al (eds) *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98,* Language Technologies Institute, Carnegie Mellon University, Pittsburgh, pp 51-61

Kaufer, D S , J R Hayes and L S Flower (1986), Composing Written Sentences, *Research in the Teaching of English,* Vol 20, pp 121-140

Kenny, Dorothy (forthcoming), Translation Units and Corpora, in Kruger, Alet (ed), *Corpus-Based Translation Studies More Research and Applications,* Manchester St Jerome

Kincaid, Calliopi (1997), A Validation Study of Simplified English as a Facilitator in English for Special Purpose Language Learning, *Technical Report* , University of Central Florida, Orlando, Florida

Knight, Kevin and Ishwar Chander (1994), Automated Postediting of Documents, in Proceedings of the 12th National Conference on Artificial Intelligence, Vol 1, Seattle, Washington, July 31-Aug 4, pp 779-784

Knops, Uus and Bart Depoortere (1998), Controlled Language and Machine Translation, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98,* Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 42-50

Kohl, J. R. (1999), Improving Translatability and Readability with Syntactic Cues, *Technical Communication*, Volume 46 (2), pp. 149-166.

Kohn, K. (1988), Fachsprache und Fachübersetzen: Psycholinguistische Dimensionen der Fachsprachenforschung, in C. Gnutzmann (ed), Fachbezogener Fremdsprachenunterricht, Tübingen: Narr, pp. 39-65.

Krings, Hans P. (2001), *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*, Kent, Ohio: The Kent State University Press, edited/translated by G.S. Koby.

Krings, Hans P (1987), The Use of Introspective Data in Translation, in C. Faerch and G. Kasper (eds), *Introspection in Second Language Research*, Clevedon: Multilingual Matters, pp.158-176.

Krings, Hans P. (1986), *Was in den Köpfen von Übersetzern vorgeht*, Tübingen: Gunter Narr Verlag.

Kussmaul, Paul and Sonja Tirkkonen-Condit (1995), Think-Aloud Protocol Analysis in Translation Studies, in *TTR*, 8, 1, pp.177-199.

Lavorel, Bernard (1982), Experience in English-French Post-Editing, in Veronica Lawson (ed), Practical Experience of Machine Translation, Proceedings of a Conference, London, 5-6 November, pp. 105-109.

Lawson, Veronica (ed) (1982), *Practical Experience of Machine Translation, Proceedings of a Conference*, London, 5-6 November 1981, Amsterdam, NewYork, Oxford: North-Holland Publishing Company.

Livbjerg, Inge and Inger M. Mees, Patterns of Dictionary Use in Non-Domain-Specific Translation, in Alves, Fabio (ed), *Triangulating Translation: Perspectives in Process Oriented Research*, Amsterdam: John Benjamins, pp. 123-136.

Loffler-Laurian, Anne-Marie (1996), *La Traduction Automatique*, Paris: Presses Universitaires du Septentrion.

Loffler-Laurian, Anne-Marie (1986a), Post-édition Rapide et Post-édition Conventionelle: Deux Modalités d'une Activité Spécifique I, in *Multilingua*, Vol. 5, Pt. 4, pp. 81-88.

Loffler-Laurian, Anne-Marie (1986b), Post-édition Rapide et Post-édition Conventionelle: Deux Modalités d'une Activité Spécifique II, in *Multilingua*, Vol. 5, Pt. 4, pp. 225-229.

Loffler-Laurian, Anne-Marie (1985), Traduction Automatique et Style, in *Babel*, Vol. 31, No. 2, pp. 70-76.

Loffler-Laurian, Anne-Marie (1984), Machine Translation: What Type of Post-Editing on What Type of Documents for What Type of Users?, in *Proceedings of the 10th International Conference on Computational Linguistics and the*

22nd Annual Meeting of the Association for Computational Linguistics - Coling '84, July 2-6, Stanford University, California, pp. 236-238.

Loffler-Laurian, Anne-Marie (1983), Pour une typologie des erreurs dans la traduction automatique, in *Multilingua*, Vol. 2, Pt. 2, pp. 65-78.

Loffler-Laurian, Anne-Marie (1981), Remarks on Some Examples of Computer-Assisted Translation and Post-Edition; Remarques à propos de quelques exemples de traduction assistée par ordinateur et la post-édition, in *Contrastes*, 2, Nov, pp. 63-69.

Lommel, Arle and Rebecca Ray (2006), *Global Business Practices: Results of the 2006 Business Practices Survey*, Localisation Industry Standards Association (LISA), Switzerland: SMP Marketing Sarl.

Lörscher, Wolfgang (1996), A Psycholinguistic Analysis of Translation Processes, in *Meta*, XLI, pp. 26-32.

Lörscher, Wolfgang (1991), *Translation Performance, Translation Process, and Translation Strategies: A Psycholinguistic Investigation*, Tübingen: Gunter Narr Verlag.

Lörscher, Wolfgang (1989), Models of the Translation Process: Claim and Reality, in *Target*, 1:1, pp. 43-68.

McElhaney, Terrence and Muriel Vasconcellos (1988), The Translator and the Postediting Experience, in Vasconcellos, Muriel (ed), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol. II, State University of New York at Binghamton (SUNY), pp. 140-148.

Means, Linda and Kurt Godden (1996), The Controlled Automotive Service Language (CASL) Project, in Adriaens et al. (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp. 106-114.

Mitamura, Teruko and Eric Nyberg (1995), Controlled English for Knowledge-Based MT: Experience with the KANT System, in *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95)*, Leuven, Belgium, July 5-7, pp. 158-172.

Mitamura, Teruko, Eric Nyberg, Geert Adriaens, Linda Schmandt, Rick Wojcik, Remi Zajac (1998) (eds), *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Mohr, S. (1985), Fehleranalyse eines maschinell übersetzten Textes im Hinblick auf Konsequenzen für Prä- und Postedition, Unpublished *Diplom* thesis, Saarbrücken: University of the Saarland, Fachrichtung Andewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen.

Mossop, Brian (2001), *Revising and Editing for Translators*, Manchester: St Jerome.

Newton, John (1992) The Perkins Experience, in Newton, John (ed), *Computers and Translation: A Practical Appraisal*, London and New York: Routledge, pp. 46-57.

Nyberg, Eric, Teruko Mitamura, Willem-Olaf Huijsen (2003), Controlled Language for Authoring and Translation, in Somers, H. (ed), *Computers and Translation: A Translator's Guide*, Amsterdam: John Benjamins, pp. 245-282.

Nyberg, Eric and Teruko Mitamura (1996), Controlled Language and Knowledge-Based Machine Translation: Principles and Practice, in Adriaens et al. (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96),* Centre for Computational Linguistics, Leuven, Belgium, pp. 74-83.

O'Brien, Sharon (2006), Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output, in *Across Languages and Cultures*, Volume 7, no. 1, pp. 1-21.

O'Brien, Sharon (2005) Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Text Translatability, *Machine Translation*, Volume 19, No. 1, pp. 37-58.

O'Brien, Sharon (2003), Controlling Controlled English - An Analysis of Several Controlled Language Rule Sets, in *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (CLAW 2003)*, 15th-17th May, Dublin City University, Ireland, Dublin: Ireland, pp. 105-114.

O'Brien, Sharon (1999), Translation Technology – The Next Generation, in *Proceedings of Translating and the Computer 21*, 10-11 November, London: Aslib (no page numbers).

O'Brien, Sharon (1998), Practical Experience of Computer-Aided Translation Tools in the Software Localisation Industry, in Bowker, Lynn, M. Cronin, D. Kenny, and J. Pearson (eds) *Unity in Diversity – Recent Trends in Translation Studies*, Manchester: St. Jerome, pp. 115-122.

Odlin, Terence (1989), *Language Transfer: Cross-Linguistic Influence in Language Learning*, Cambridge: Cambridge University Press.

Ogden, Charles Kay (1930), *Basic English: A General Introduction with Rules and Grammar*, London: Paul Treber & Co. Ltd.

Olohan, Maeve (ed) (2000), *Intercultural Faultlines: Research Models in Translation Studies: Textual and Cognitive Aspects*, Manchester: St. Jerome.

Olohan, Maeve (1991), *An Introspection-Based Analysis of the Post-Editing Process*, Unpublished M A thesis, Dublin Dublin City University

Pigott, Ian M (1988), MT in Large Organizations Systran at the Commission of the European Communities, in Vasconcellos, Muriel (ed) *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol II, State University of New York at Binghamton (SUNY), pp 159-166

Pigott, Ian M (1982), The Importance of Feedback from Translators in the Development of High-Quality Machine Translation, in Veronica Lawson (ed) Practical Experience of Machine Translation, Proceedings of a Conference, London, 5-6 November, pp 61-73

Povlsen, Claus and Annelise Bech (2002), Ape Reducing the Monkey Business in Post-Editing by Automating the Task Intelligently, in *Proceedings of MT Summit VIII, Machine Translation in the Information Age*, Santiago de Compostela, Spain,18-22 September, pp 283-286

Price, Jonathan (1984), *How to Write a Computer Manual – A Handbook of Software Documentation*, California Benjamin/Cummings Publishing Company

Reuther, Ursula (2003), Two in One- Can it Work? Readability and Translatability by means of Controlled Language, in *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (CLAW 2003)*, 15th-17th May, Dublin City University, Ireland, pp 124-132

Reuther, Ursula (1998), Controlling Language in an Industrial Application, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 174-184

Reuther, Ursula and Antje Schmidt-Wigger (2000), Designing a Multi-Purpose CL Application, in Adriaens et al (eds), *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*, Seattle, Washington, pp 72-82

Robertson, F A and Edward Johnson (1988), *Airspeak Radiotelephony Communication for Pilots*, Oxford Prentice Hall

Rossi, Francesco (1982), The Impact of Posteditors' Feedback on the Quality of MT, in Veronica Lawson (ed), *Practical Experience of Machine Translation, Proceedings of a Conference*, London, 5-6 November, pp 113-118

Rothe-Neves, Rui (2003), The Influence of Working Memory Features on Some Formal Aspects of Translation Performance, in Fabio Alves (ed), *Triangulating Translation Perspectives in Process Oriented Research*, Amsterdam John Benjamins, pp 97-119

Roturier, Johann (2004), Assessing a Set of Controlled Language Rules Can They Improve the Performance of Commercial Machine Translation Systems?, in *Proceedings of Translating and the Computer 26, Aslib 2004*, London Aslib (no page numbers)

Ryan, Joann P (1988), The Role of the Translator in Making an MT System Work Perspective of a Developer, in Vasconcellos, Muriel (ed), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol II, State University of New York at Binghamton (SUNY), pp 127-132

Sågvall Hein, Anna (1997), Language Control and Machine Translation, in *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in MT (TMI-97)*, Santa Fe, USA, (no page numbers)

Sågvall Hein, Anna, Ingrid Almqvist, Per Starback (1997), ScaniaSwedish - A Basis for Multilingual Machine Translation, In *Proceedings of Translating and the Computer 19 - ASLIB* 97, November, London, (no page numbers)

Sågvall Hein, Anna, Ingrid Almqvist (2000), A Language Checker of Controlled Language and its Integration into a Documentation and Translation Workflow, in *Proceedings of Translating and the Computer 22 - ASLIB 2000*, London, November, (no page numbers)

Sander, B (1990), *Untersuchung zur Postedition maschinell ubersetzter Texte - Sprachrichtung Deutsch-Englisch*, Unpublished *Diplom* thesis, Hildesheim University of Hildesheim, Institut fur Angewandte Sprachwissenschaft

Sandrock, U (1982), *Thinking-Aloud Protocols (TAPs) - Ein Instrument zur Dekomposition des komplexen Prozesses Ubersetzen*, Unpublished Staatsarbeit, Kassel University of Kassel

Santangelo, Susana (1988), Making an MT System Work Perspective of a Translator, in Vasconcellos, Muriel (ed), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol II, State University of New York at Binghamton (SUNY), pp 133-139

Sayans Gomez, Antonio and Elena Villar Conde (2003), The Functionality of a Toolbar for Postedition in Machine Translation between Languages with Linguistic Interference the Spanish Galician Case, in *Online Proceedings of MT Summit IX*, New Orleans, http //www amtaweb org/summit/MTSummit/papers html [Last accessed 06/12/2005]

Schachtl, Stefanie (1996), Requirements for Controlled German in Industrial Applications, in Adriaens et al (eds) *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96)*, Centre for Computational Linguistics, Leuven, Belgium, pp 143-149

Schilperoord, J (1996), It's About Time Temporal Aspects of Cognitive Processes in Text Production, Amsterdam Rodopi

Schmidt-Wigger, Antje (1998), Grammar and Style Checking for German, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 76-86

Schreurs, Dirk and Geert Adriaens (1992), From COGRAM to ALCOGRAM Toward a Controlled English Grammar Checker, in *Proceedings of the 14th International Conference on Computational Linguistics - COLING '92*, pp 595-601

Schutz, Jorg (2001), Controlled Language Deployment New Challenges and Opportunities for Translation Professionals, in *Proceedings of the Federcentri Conference*, Bologna, 26th-28th October, pp 191-210

Schwertel, Uta (2000), Controlling Plural Ambiguities in Attempto Controlled English (ACE), in Adriaens et al (eds), *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*, Seattle, Washington, pp 105-119

Schwitter, Rolf, A Ljungberg, D Hood (2003), ECOLE A Look-Ahead Editor for a Controlled Language, in *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (CLAW 2003)*, 15th-17th May, Dublin City University, Ireland, pp 141-150

Schwitter, Rolf and Norbert Fuchs (1996), Attempto - From Specifications in Controlled Natural Language towards Executable Specifications, EMISA Workshop *Naturlicher Entwurf von Informationssystemen*, Akademie Tutzing, May 28-30, (no page numbers)

Séguinot, Candace (1997), Accounting for Variability in Translation, in Danks, Joseph, Gregory Shreve, Stephen Fountain, and Michael McBeath (eds), *Cognitive Processes in Translation and Interpreting*, Thousand Oaks, California Sage Publications, pp 104-119

Séguinot, Candace (ed) (1989a), *The Translation Process*, Toronto H G Publications, School of Translation, York University

Seguinot, Candace (1989b), The Translation Process An Experimental Study, in *The Translation Process*, Toronto H G Publications, School of Translation, York University, pp 21-53

Senez, Dorothy (1998a), Post-Editing Service for Machine Translation Users at the European Commission, in *Proceedings of Translating and the Computer 20*, London Aslib, (no page numbers)

Senez, Dorothy (1998b) The Machine Translation Help Desk and the Post-Editing Service, in *Terminologie et Traduction la Revue des Services Linguistiques des Institutions Européennes*, Vol 1, pp 289-295

Sereda, Stanley (1982), Practical Experience of Machine Translation, in Veronica Lawson (ed), *Practical Experience of Machine Translation, Proceedings of a Conference*, London, 5-6 November, pp 119-123

Shubert, Serena, Jan Spyridakis, Heather Holmback, M B Coney (1995), The Comprehensibility of Simplified English in Procedures, in *Journal of Technical Writing and Communication*, 25 (4), pp 347-369

Smart, John (1988), Getting Smart in Many Languages MT with an Option of Preprocessing, in Muriel Vasconcellos (ed), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol II, State University of New York at Binghamton (SUNY), pp 124-126

Somers, Harold (1997), A Practical Approach to Using MT Software - "Post-editing" the Source Text, in *The Translator*, 3, 2, Nov, pp 193-212

Spyridakis, Jan, Serena Shubert, Heather Holmback (1997), Measuring the Translatability of Simplified English Procedures, *IEEE Transactions on Professional Communication*, 40, 1, pp 217-246

TAUS, Translation Automation User Society, http //www translationautomation com [Last accessed 29/03/2006]

Tirkkonen-Condit, Sonja (ed) (1991), *Empirical Research in Translation and Intercultural Studies, Selected Papers of the TRANSIF Seminar, Savonlinna 1988*, Tubingen Gunter Narr Verlag

Tirkkonen-Condit, Sonja (1989), Professional vs Non-Professional Translation A Think-Aloud Protocol Study, in Séguinot C (ed), *The Translation Process*, Toronto H G Publications, School of Translation, York University pp 73-85

Toury, Gideon (1991), Experimentation in Translation Studies Achievements, Prospects and Some Pitfalls, in Tirkkonen-Condit, S (ed), *Empirical Research in Translation and Intercultural Studies, Selected Papers of the TRANSIF Seminar, Savonlinna 1988*, Tubingen Gunter Narr Verlag, pp 89-109

Trujillo, Arturo (1999), *Translation Engines. Techniques for Machine Translation*, London Springer-Verlag

Tyson, Denis W F (1985), Planning and Commissioning Technical Translations, in Austin, M (ed), *The ISTC Handbook of Technical Writing and Publication Techniques*, London Heinemann, pp 119-130

Underwood, Nancy and B Jongejan (2001), Translatability Checker A Tool to Help Decide Whether to Use MT, in Maegaard, B (ed), *Proceedings of the MT Summit VIII Machine Translation in the Information Age*, 18-22 September, Santiago de Compostela, Spain, pp 363-368

Van Der Meer, Jaap (2003), At Last Translation Automation Becomes a Reality An Anthology of the Translation Market, in *Proceedings of the Joint*

Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop (CLAW 2003), 15th-17th May, Dublin City University, Ireland, pp 180-184

Van Slype, Georges (1982), Economic Aspects of Machine Translation, in Veronica Lawson (ed), *Practical Experience of Machine Translation, Proceedings of a Conference*, London, 5-6 November, pp 79-93

Van Waes, Luuk and P J Schellens (2003), Writing Profiles The Effect of the Writing Mode on Pausing and Revision Patterns of Experienced Writers *Journal of Pragmatics*, vol 35, pp 829-853

Vasconcellos, Muriel (ed) (1988a), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol II, State University of New York at Binghamton (SUNY)

Vasconcellos, Muriel (1986a), Post-editing On-Screen Machine Translation from Spanish into English, in *Proceedings of Translating and the Computer 8*, London Aslib, November, (no page numbers)

Vasconcellos, Muriel (1986b), Functional Considerations in the Postediting of Machine Translated Output, Dealing with V(S)O versus SVO, in *Computers and Translation*, 1 1, pp 21-38

Vasconcellos, Muriel (1985), Management of the Machine Translation Environment Interaction of Functions at the Pan American Health Organization, in Veronica Lawson (ed), *Tools for the Trade, Translating and the Computer 5*, London Aslib, pp 115-129

Vasconcellos, Muriel and Marjorie León (1985), SPANAM and ENGSPAN Machine Translation at the Pan American Health Organization, *Computational Linguistics*, Vol 11, No 2-3, pp 122-136

Veale, Tony and Andy Way (1997), Gaijin A Bootstrapping Approach to Example-Based Machine Translation, in *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing* , 01-05 September, Tzigov Chark, Bulgaria (no page numbers)

Wagner, Elizabeth (1985), Rapid Post-Editing of Systran, in Veronica Lawson (ed), *Tools for the Trade Proceedings of Translating and the Computer 5*, pp 199-213

Wagner, Emma (1987), Post-editing - Practical Considerations, in Catriona Picken (ed), *ITI Conference I The Business of Translating and Interpreting*, London Aslib, pp 71-78

Wagner, Emma (1985), Post-editing Systran - A Challenge for Commission Translators, *Terminologie et Traduction*, 3-1985, OPOCE, European Commission

Wells-Akis, Jennifer, and W R Sisson (2002), Improving Translatability - A Case Study at Sun Microsystems Inc , *The LISA Newsletter Globalization Insider*, No 4 5

White, E N (1980), Using Controlled Languages for Effective Communication, in *Proceedings of the 27th International Technical Communication Conference*, Society for Technical Communication, Washington DC, pp E110-E113

Wilms, F J (1981), Von Susy zu Susy-BSA - Forderungen an ein anwenderbezogenes MU-System, *Sprache und Datenverarbeitung*, I(2), pp 38-43

Wojcik, Rick, Heather Holmback, James Hoard (1998), Boeing Technical English An Extension of AECMA SE beyond the Aircraft Maintenance Domain, in Mitamura et al (eds), *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, pp 114-123

Wojcik, Rick and James Hoard (1996), Controlled Languages in Industry, in Cole, Ronald et al (eds), *Survey of the State of the Art in Human Language Technology*, http //cslu cse ogi edu/HLTsurvey/HLTsurvey html [Last Accessed 23/03/2006]

Wojcik, Rick and Heather Holmback (1996), Getting a Controlled Language off the Ground at Boeing, in Adriaens et al (eds), *Proceedings of the First International Workshop on Controlled Language Applications (CLAW 96)*, Centre for Computational Linguistics, Leuven, Belgium, pp 22-31

Woyde, Rick (2001), Introduction to the SAE J2450 Translation Quality Metric, in *Language International*, April, pp 37-39

Zabalbeascoa, Patrick (2000), From Techniques to Types of Solutions, in Beeby, Allison, Doris Ensinger and Maria Presas (eds), Investigating Translation Selected Papers from the 4th International Congress on Translation, Barcelona 1998, Amsterdam/Philadelphia John Benjamins, pp 117-127

Zucko, D (1993), *Der Einfluß der Vorredaktion auf den Nachredaktionsaufwand bei Maschinenubersetzungen - Eine empirische Untersuchung*, Unpublished *Diplom* thesis, Hildesheim University of Hildesheim, Institut fur Angewandte Sprachwissenschaft