DUBLIN CITY UNIVERSITY

Community Post-Editing of Machine-Translated User-Generated Content

Author:
Linda MITCHELL, B.A., M.A.

Supervisors:
Dr. Sharon O'BRIEN
Dr. Johann ROTURIER
Dr. Fred HOLLOWOOD

Thesis submitted for the degree of the Doctor of Philosophy

in the

School of Applied Language and Intercultural Studies

3rd February 2015

External Examiner:
Dr. Joanna DRUGAN
University of East Anglia

Internal Examiner:
Dr. Dorothy KENNY

Declaration of Authorship

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 that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____   ID No: _____

Date: _____

Acknowledgements

Publications & Presentations Arising from this Research

Publications from this Research Project:

Mitchell, L., O'Brien, S. and Roturier, J. 2014. Quality Assessment in Community Post-Editing. Machine Translation: Special Issue on Post-editing 28(3-4), pp. 237-262.

Mitchell, L., Roturier J. and Silva, D. 2014. Using the ACCEPT framework to conduct an online community-based translation evaluation study. In: *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 16-18 June*, Dubrovnik, Croatia, pp. 155-158.

Mitchell, L., Roturier, J. and O'Brien, S. 2013. Community-based post-editing of machine-translated content: monolingual vs. bilingual. In: *Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2013)*, 2-6 September, Nice, France, pp. 35-44.

Roturier, J., Mitchell, L. and Silva, D. 2013. The ACCEPT Post-Editing Environment: a Flexible and Customisable Online Tool to Perform and Analyse Machine Translation Post-Editing. In: *Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2013)*, 2-6 September, Nice, France, pp.119-128.

Mitchell, L. and Roturier, J. 2012. Evaluation of machine-translated user generated content: a pilot study based on user ratings. In: *EAMT 2012: Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, Trento, Italy, pp.61-64.

Presentations from this Research Project:

Evaluating Community Post-Editing - Bridging the Gap between Translation Studies and Social Informatics. *Second Durham Postgraduate Colloquium: Theoretical Frameworks and Methodologies: Research in Translation Studies*. Durham University, UK, 31 January 2014.

Monolingual Post-Editing in an Online Community. *International Association for Intercultural Studies (IATIS) Conference*. Queens University, Belfast, 24 July 2012.

Evaluating community post-edited content. *TAUS Quality Evaluation Summit held at Localization World*. Dublin, Ireland, 4 June 2014.

Publications from ACCEPT :

Roturier, J., Mitchell, L., Grabowski, R. and Siegel, M. 2012. Using Automatic Machine Translation Metrics to Analyze the Impact of Source Reformulations. In: *Proceedings of the 10th biennial conference of the Association for Machine Translation in the Americas (AMTA-2012), 28 October - 1 November*, San Diego, USA, pp.10.

Presentations from ACCEPT:

ACCEPT - Automated Community Content Editing PorTal, *Machine Translation Summit XIV: Projects Village, 4–6 September 2013*, Nice, France.

ACCEPT - Automated Community Content Editing PorTal. *Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT), 28-30 May*, Trento, Italy.

Contents

Abbreviations

ACCEPT	**A**utomated **C**ommunity **C**ontent **E**diting **P**or**T**al
AEM	**A**utomatic **E**valuation **M**etric
ALPAC	**A**utomatic **L**anguage **P**rocessing **A**dvisory **C**ommittee
AMT	**A**mazon **M**echanical **T**urk
AMTA	**A**ssociation for **M**achine **T**ranslation in the **A**mericas
API	**A**pplication **P**rogramming **I**nterface
BLAST	**B**i**L**ingual **A**nnotation/Annotator/Analysis **S**upport **T**ool
BLEU	**B**i**L**ingual **E**valuation **U**nderstudy
BT	**B**ack **T**ranslation
CASMACAT	**C**ognitive **A**nalysis and **S**tatistical **M**ethods for **A**dvanced
	Methods for **A**dvanced **C**omputer **A**ided **T**ranslation
CAT	**C**omputer-**A**ided **T**ranslation
CoSyne	**C**ontent **Syn(e)**chronization
CL	**C**ontrolled **L**anguage
CNA	**C**hoice **N**etwork **A**nalysis
CNGL	formerly known as: **C**entre for **N**ext **G**eneration **L**ocalisation,
	now: Centre for Global Intelligent Content
CPE	**C**rowd **P**ost-**E**diting
DST	**D**ynamic **S**ystems **T**heory
DQF	**D**ynamic **Q**uality **F**ramework
EUROPARL	**EURO**pean **PARL**iament
FSTS	**F**ormalized **S**tructured **T**ranslation **S**pecifications
GALE	**G**lobal **A**utonomous **L**anguage **E**xploitation
HT	**H**uman **T**ranslation
ISLC	**I**nformation **S**ystems **L**ife **C**ycle

IT	**I**nformation **T**echnology
ITP	**I**nteractive **T**ranslation **P**rediction
JSON	**J**ava**S**cript **O**bject **N**otation
LDC	**L**inguistic **D**ata **C**onsortium
LISA	**L**ocalization **I**ndustry **S**tandards **A**ssociation
LM	**L**anguage **M**odel
LSP	**L**anguage **S**ervice **P**rovider
MT	**M**achine **T**ranslation
NIST	**N**ational **I**nstitute of **S**tandards and **T**echnology
NTI	**N**egative **T**ranslatability **I**ndicator
OC	**O**nline **C**ommunity
PACTE	**P**rocés d'**A**dquisició de la **C**ompetència **T**raductora i Avluació
PE	**P**ost-**E**diting
PET	**P**ost-**E**diting **T**ool
PU	**P**roduction **U**nit
QA	**Q**uality **A**ssurance
QE	**Q**uality **E**valuation
RQ	**R**esearch **Q**uestion
SL	**S**ource **L**anguage
SPSS	**S**tastical **P**ackage for the **S**ocial **S**ciences
ST	**S**ource **T**ext
SMT	**S**tatistical **M**achine **T**ranslation
TAP	**T**hink-**A**loud **P**rotocol
TAUS	**T**ranslation **A**utomation **U**ser **S**ociety
TC	**T**ranslation **C**ompetence
TL	**T**arget **L**anguage
TM	**T**ranslation **M**emory
TT	**T**arget **T**ext
TER	**T**ranslation **E**rror **R**ate
UGC	**U**ser-**G**enerated **C**ontent
UI	**U**ser **I**nterface
WELL	**W**hole **E**arth 'Lectronic **L**ink
WER	**W**ord **E**rror **R**ate

WMT	**W**orkshop on Statistical **M**achine **T**ranslation
WP	**W**ork **P**ackage
XLIFF	**X**ML **L**ocalisation **I**nterchange **F**ile **F**ormat

List of Figures

List of Tables

Abstract

Community Post-Editing of Machine-Translated User-Generated Content,
by Linda MITCHELL

With the constant growth of user-generated content (UGC) online, the demand for quick translations of large volumes of texts increases. This is often met with a combination of machine translation (MT) and post-editing (PE). Despite extensive research in post-editing with professional translators or translation students, there are few PE studies with lay post-editors, such as domain experts. This thesis explores lay post-editing as a feasible solution for UGC in a technology support forum, machine translated from English into German. This context of lay post-editing in an online community prompts for a redefinition of quality.

We adopt a mixed-methods approach, investigating PE quality quantitatively with an error annotation, a domain specialist evaluation and an end-user evaluation. We further explore post-editing behaviour, i.e. specific edits performed, from a qualitative perspective. With the involvement of community members, the need for a PE competence model becomes even more pressing. We investigate whether Göpferich's translation competence (TC) model (2009) may serve as a basis for lay post-editing.

Our quantitative data proves with statistical significance that lay post-editing is a feasible concept, producing variable output, however. On a qualitative level, post-editing is successful for short segments requiring $\leq 35\%$ post-editing effort. No post-editing patterns were detected for segments requiring more PE effort. Lastly, our data suggests that PE quality is largely independent of the profile characteristics measured.

This thesis constitutes an important advance in lay post-editing and benchmarking the evaluation of its output, uncovering difficulties in pinpointing reasons for variance in the resulting quality.

Chapter 1

Introduction

1.1 Research Background

The funding for this thesis was awarded by Symantec Corporation. The supervision was jointly conducted by Symantec and by the School of Applied Language and Intercultural Studies, Dublin City University (DCU).

Under Symantec's affiliation, this thesis constitutes a part of the ACCEPT project (Automated Community Content Editing PorTal),[1] funded by the Seventh Framework Programme (Grant Agreement Number: 288769). The ACCEPT project aimed at connecting speakers of different languages online by developing an intuitive, unintrusive and dynamic translation solution for both commercial and non-profit purposes. The ACCEPT project focussed on creating technology that supports the automatic translation of text with three interventions. Firstly, it developed language-specific pre-editing rules to eliminate errors in the source text (ST) to be translated; secondly, it developed a statistical machine translation (SMT) system based on novel domain

---

[1]The project timeline was from 1 January 2012 - 31 December 2014.

adaptation methods; thirdly, it developed language-specific post-editing (PE) rules to correct errors that remain in the machine translated (MT) output. To evaluate machine translated/post-edited output, it fourthly developed an evaluation functionality. In order to facilitate a holistic user experience, the ACCEPT project created a web portal[2] that integrates all four functionalities. In addition, open-source post-editing and pre-editing plug-ins and the option of integrating and evaluating translation projects into any platform were created. This PhD thesis mainly focussed on WP7 (Work Package) and WP8 of the ACCEPT project, dealing with monolingual and bilingual post-editing. It provided a subset of the research objectives within the ACCEPT project by investigating the feasibility of PE in the German Norton Community. Privileged access to the community was provided by the project partner Symantec, while the pre-editing, machine translation and post-editing technologies employed were developed and provided by the ACCEPT Consortium.

1.2   The Norton Community

Symantec Corporation produces a number of consumer software products in the security and backup domains, e.g. the Norton product line. The Norton Business Unit is a division of Symantec that is specialised in producing software, such as antivirus protection and backup (Norton Security). From 2008 to September 2014, five separate language-specific Norton Communities existed (English, French, German, Japanese and Chinese). They were developed as a platform for customers to communicate with other customers in their native language. Each community consists of boards, which group

---

[2]The ACCEPT portal is currently hosted on: www.accept-portal.eu [Accessed: 25 November 2014]
The code is open-source and available on github: https://github.com/accept-project/accept-portal

together posts on a particular topic, e.g. the "Norton Internet Security/ Norton Anti-Virus" board solely contains threads on the "Norton Internet Security" and "Norton AntiVirus" products. Each thread typically consists of a question posted by a community member and replies from other community members.[3] One of these replies is then marked by the original poster as an "accepted solution". Apart from interacting through forum posts, customers can communicate via private messages and are able to reward posts (questions, answers, blog posts) with "kudos". These communities were selected as they firstly, were easily accessible through the sponsor of this thesis, Symantec, and, secondly, as they are communities with a primary purpose other than translation or social activity (e.g. Facebook, Twitter). Hence, it provides a novel platform for exploring post-editing online. This thesis focusses on the English-speaking and the German-speaking Norton Communities. The English-speaking forum provided the source texts to be machine translated and post-edited, while the post-editors and the community evaluators were members of the German Norton Community. In September 2014, these separate communities were migrated to a new platform[4] creating a large multi-lingual community. This, however, did not eliminate language barriers for monolingual users. The language pair English $\rightarrow$ German that was chosen deliberately, as it is a language pair for which MT engines struggle to translate content correctly, especially with added difficulties that arise from, for example, user-generated content. Hence, a form of improving the MT output, here post-editing, would be beneficial if not necessary in a real-life scenario, such as the MT content in the Norton Communities.

---

[3]Threads can also be started by announcements or commentaries, which are posted in specific boards or blogs.

[4]http://community.norton.com/

1.3 Motivation

Over the last decades, the internet has become a great knowledge base, which now serves as a primary resource to answer any type of question. Search online has been accelerated since the rise of web-enabled mobile phones and tablets, producing answers in a few clicks. The answer is also often embedded in user-generated content (UGC), such as that posted in online communities, which may be written in any language. The motivation for this research lies in the desire to increase the visibility of this type of content by eliminating language barriers online for monolingual users to have access to knowledge originally written in any language.

Solutions for overcoming online language barriers include human translation (HT) and machine translation. With increasing volumes of content to be translated, companies are struggling to keep up with translation demand. In the case of translating UGC in online communities, Kelly (2009) fittingly describes the dilemma faced: conventional human translation cannot be an efficient solution because it creates a time-delay. Before the process of translation and quality assurance can be completed, the source text in the community may already have been edited or expanded, i.e. more information may have been added to the thread. To circumvent this delay, companies are tapping into machine translation as a potential solution. This research investigates the use of machine translation and post-editing online, by embedding these technologies into the community environment. Machine translation continues to face challenges in producing high-quality output, i.e. transferring the meaning correctly using adequate language. This holds true especially for the context of translating user-generated content for language pairs that are not closely related to each other, such as MT from

English into German. Besides "noisy" input, i.e. source texts riddled with spelling and grammar mistakes, the syntactical differences between the two languages hamper the production of a correct translation. This can lead to output that is distorted, sometimes rendering the original meaning unrecognisable. The chosen intervention for addressing those difficulties is post-editing, i.e. fixing errors in the machine translated output. Currently, post-editing is typically done by translators or translation students (see Chapter 2, section 2.4). This thesis aims at exploring whether a wider pool of post-editors, in particular any member of the Norton Community, can successfully operate as lay post-editor.[5] So far, the ability of community members to successfully post-edit remains largely unresearched. This constitutes the central novel element in the research presented here and is motivated by the constant growth of communities and communication online. The term 'lay post-editor' was identified as the most appropriate and inoffensive descriptor and was derived from terminology presented in recent research in the field (cf. 'lay user' (Aranberri et al. 2014) and 'lay community' (Laurenzi et al. 2013). While lay post-editors may not possess multilingual or translation skills, they do possess knowledge relevant to the domain that may be helpful or even crucial in facilitating successful post-editing. To date, there has been little discussion of formal competence models for post-editing. Any such discussion is usually focussed on post-editing competence relative to formal translation competence: to the best of our knowledge no consideration has been given to a post-editing competence model for lay post-editors. Another novel aspect of this research, therefore, is the consideration of the relevance of an existing translation competence (TC) model for lay post-editors.

---

[5]A lay post-editor includes anybody who is not a professional translator or translation student. A professional translator is anybody who has completed translation training and is working as a translator professionally on a daily basis. Due to the anonymity of the participants, we cannot eliminate the possibility that the set of lay post-editors in the online community included professional translators.

This thesis was motivated by a need for a feasible and effective solution for translating and post-editing UGC as well as the need for a theoretical approach to investigate lay post-editor competences. An investigation of competence necessitates the consideration of quality. We identified a need for assessing quality in a manner that is appropriate to its context, i.e. in an online community context.

1.4   Research Questions

In light of the motivation for this thesis, the following four research questions were formulated:

RQ1) *How does post-edited output produced by community post-editors compare to raw machine translated output?*

RQ2) *What are segment characteristics and post-editing patterns that render post-editing successful/unsuccessful?*

RQ3) *Do elements of the post-editor profile serve as predictors for the quality of the post-edited output?*

RQ4) *Is Göpferich's Translation Competence model applicable to community post-editing?*

While the first two questions are concerned with the quality of the post-edited output, the third question focusses on the post-editor profile, which forms the basis for testing an existing Translation Competence model in the final question. These are discussed in more detail in Chapter 4, section 4.4.1.

1.5   Thesis Overview

This thesis is organised as follows. Chapters 2 and 3 constitute the literature review, which is sub-divided thematically: Part I introduces the concept of post-editing - what it is, where it came from and where it is going – and discusses translation competence models and how they might relate to post-editing in a community context. Part II presents the concept of MT and PE quality evaluation, their development and approaches commonly taken. It further investigates the topic of online communities, their development in general and their use as a platform for MT and PE.

Subsequently, Chapter 4 outlines and justifies the mixed-method research design chosen for this thesis and discusses the methods to operationalise each of the research questions, based on observations made during a pilot study. This includes the preparation for the experiment, such as post-editor recruitment, content selection and pre-processing; the experiment design, with emphasis on the editing environment and the data collection method, and the evaluation of the content, with particular focus on different views and aspects of quality based on previous experience within the fields of translation and online communities. It outlines a three-fold evaluation step involving an error annotation, a domain specialist evaluation and a community evaluation with the end-users of the translations.

Chapters 5, 6 and 7 present the results from the data collected. The presentation of the results is also split thematically; Chapter 5 addresses RQ1, the quality evaluation of the post-edited data from a quantitative perspective, while Chapter 6 addresses the PE quality from a qualitative perspective, focussing on RQ2. Chapter 7 synthesises the data collected on the post-editor profile to provide answers for RQ3 and to explore

the relation between PE quality and the post-editor profile. Simultaneously, the data collected on the post-editor profile is used to test some aspects of Göpferich's TC model, which constitutes RQ4.

Chapter 7 reflects on the initial research aims and how these have been met with the experimental design and results presented: it outlines the contributions to the field, the strengths and shortcomings of the methodological design and potential directions for future research.

Chapter 2

Literature Review - Part I

2.1 Introduction

This and the following chapter constitute the literature review for this thesis. Chapter 2 focusses on the discussion of a theoretical framework for post-editing and relevant research in post-editing. Chapter 3 discusses the evaluation of machine translated and more importantly post-edited content followed by presenting online communities as a platform for research in post-editing.

Firstly, this chapter establishes a theoretical framework for post-editing research, prompted by a lack of coverage of a theoretical framework for research in PE and by a need to understand what skills successful post-editors should possess. For this purpose, translation competence models that have been developed for traditional approaches to translation (PACTE 2005, Göpferich 2013) are investigated and discussed

with their relevance to lay translators performing post-editing in a collaborative manner online. Based on this, a strategy is outlined for taking a step towards the development of a lay post-editing competence model. Secondly, this chapter explores the need for post-editing and the different types of post-editing. It introduces early studies but emphasises more recent ones, focussing on cognition, quality, effort and PE tools. Subsequently, this chapter investigates post-editing from a crowd-sourcing perspective, prompted by online communities becoming more popular as conduits for translation. Looking at previous research, it distinguishes between crowd post-editing and community post-editing, the latter of which is the central subject of study here.

Finally, it discusses the relevance of previous research to this context and outlines how this thesis seeks to extend the body of existing knowledge.

2.2 Theoretical Framework

2.2.1 Translation Competence (TC)

This thesis focusses on the machine translation and post-editing of content of a technical domain mixed with colloquial and general language in the medium of user-generated content. The main purpose of this content that resides in the Norton Communities is to explain the steps that are needed to solve a particular problem accurately. For this purpose, the meaning conveyed of these texts takes the highest priority. For this thesis then, translation involves transferring the meaning from a source text (ST) of one language to a target text (TT) in another language. As is the case for human translation, the post-editor seeks to transfer the same meaning from one text to the other. Post-editing, however, involves "checking, proof-reading

and revising translations carried out by any kind of translation automaton" Gouadec (2007). One could thus argue that translation competence is a determining factor for successful post-editing. Currently, neither a post-editing competence model in general, nor a post-editing (or translation competence) model for lay translators, exists. Skills that should be taught in relation to post-editing were based on translation skills, as proposed by O'Brien (2002), though there is limited evidence that PE is currently being taught in university translator training programmes. Due to the lack of a post-editing competence model, a translation competence model forms the starting point for a theoretical basis here. From a translation perspective, it has been deemed inconceivable that lay translators should be, or are, able to perform translation or post-editing tasks successfully (e.g. Pierce et al. 1966, p. 97). This sentiment is also reflected in the existing TC models, which were developed for the training of professional translators and their acquisition of translation competence only.

The development of a theoretical model for Translation Competence was reported in the 1970s/1980s by translation scholars such as Harris (1973, 1976) and Gideon Toury (1980, 1995, 2012). Both scholars focus on an initial natural state of translation, without formal translation training, i.e. that of "natural translation" (Harris 1976) and a "native translator" (Toury 1980). Toury argues that bilingualism is neither completely separable from translator competence, nor is it the only prerequisite for translation competence. Toury (2012) calls into question whether translation can be seen as humanly innate of bilinguals, as proposed by Harris and Sherwood (1978). These two arguments are of importance to this thesis, as it remains to be seen whether they prove to be true in a post-editing context with lay translators.

2.2.1.1   TC according to PACTE

Since then, more comprehensive theoretical models for translation competence and translation competence acquisition have been developed. The PACTE Group has been carrying out research in the field of translation competence for some years now, focussing both on the translation process on a cognitive level and the translation product on a textual level. After proposing and testing their TC model in the early years (PACTE 2002, PACTE 2005), they report results from an extensive study with 35 translators (trained and working in translation) and 24 lay translators (foreign-language teachers) in PACTE (2011). They argue that translation competence is comprised of the following five sub-competences: 1) Bilingual sub-competence, 2) Extra-linguistic sub-competence, 3) Knowledge about translation, 4) Instrumental sub-competence which involves "procedural knowledge related to the use of documentation resources and information and communication technologies applied to translation" (2011, p. 319) and 5) Strategic sub-competence, comprised of "procedural knowledge to guarantee the efficiency of the translation process and solve problems encountered [...] [and] serves to control the translation process" (2011, p. 319). PACTE regards the aspect of 6) Psycho-physiological components as integral but not specific to translation, which is why it is not included as a sub-competence.

The PACTE Group believes that all bilinguals possess the first two sub-competences, which constitutes the reason for their focussing specifically on the remaining sub-competences. In particular, they investigate translation problems, which they define as "when "automised" solutions, i.e. spontaneous and immediate solutions, are not found for source-text segments in translation and different strategies are then put into effect to solve them" (PACTE 2009, p. 328), and the ability to solve the same.

They seek to answer questions focussing on investigating TC of translators vs. lay translators:

- Is the ability to identify and solve translation problems really a characteristic of translation competence?
- Do translators and non-translators[1] behave in the same way?
- Do they perceive the same degree of difficulty in translating a text?
- Do they identify the same kinds of problem?
- Do they describe the problems in the same terms? (PACTE 2009, p. 329).

The PACTE Group attempts to establish whether there is a difference between TC for professional translators and lay translators. They employ self-reporting in surveys to measure translation problems and to investigate the translation process. They find that both groups have a different perception of the level of difficulty of translation depending on the language pair. They do not report statistical significance but use the term "relevant" difference. They find that all participants struggled to verbalise the difficulties they encountered during translation and state that there is no notable difference between translators and lay translators in their characterisation of translation problems. They also report a lack of correlation between the perceived acceptability of a translation by the participant and acceptability as measured by a professional evaluator. Besides the possibility that the concepts are not correlated, this may suggest that self-reporting is not an appropriate method of rating the difficulty and the acceptability of a translation.

Returning to the TC model, from the homogeneous data gathered on describing translation problems encountered by both translators and lay translators one may conclude that the ability to identify and describe translation problems is independent of translator training. Based on this, lay translators may be able to perform translation

---

[1] PACTE define translators as expert translators and non-translators as theoretically everybody else but in their experiment they are foreign-language teachers.

tasks or post-editing tasks, provided that they can compensate for a lack in transla-
tion knowledge or language skills through, for example, extra-linguistic skills (domain
competence) and/or strategic competence. Going forward, it would be beneficial to
establish, whether translation training explicitly included the discussion of translation
problems and if so to what extent.

2.2.1.2 TC according to Göpferich

Göpferich (2013) gathers comprehensive data on the translation process and thus the
development of translation competence by conducting a longitudinal study (TransComp),
investigating the development of translation competence in 12 translation students over
the course of three years. Prior to TransComp, she developed a TC model (Göpferich
2009a), reviewing other strategies of approaching translation competence (Göpferich
2013). Göpferich compares the ideal translation process developed by Hönig (1991,
1995) to the minimalist approach suggested by Pym 2003, which share two main sub-
competences, associative competence and the competence to develop a macro-strategy
and to apply it consistently. She compares these to the TC model developed by the
PACTE Group (e.g. 2003). Göpferich takes this model as a basis, modifies and adds
to it to suit her purposes for the TransComp project. She argues that her approach
combines relevant aspects of the approaches presented above. She carries out changes,
which she regards essential in order to make the model more accessible and more apt.

In particular, she adopts four of the sub-competences identified by PACTE in her
model, and reduces the aspect of knowledge about translation sub-competence to

translation routine activation competence to make it more tangible. She also extracts psycho-motor competence from PACTE's psycho-physiological component and integrates it as a competence into the model.

The Göpferich model thus consists of the components displayed in Figure 2.1. On the basis of Göpferich's argumentation (cf. Göpferich 2009a, pp. 13-21) her model was identified as the most suitable as a starting point for an exploration into identifying a (lay) post-editing competence model. While this dissertation focusses on the central competence contained in the circle, she includes additional factors outside these competences, such as translation brief and translation norms.
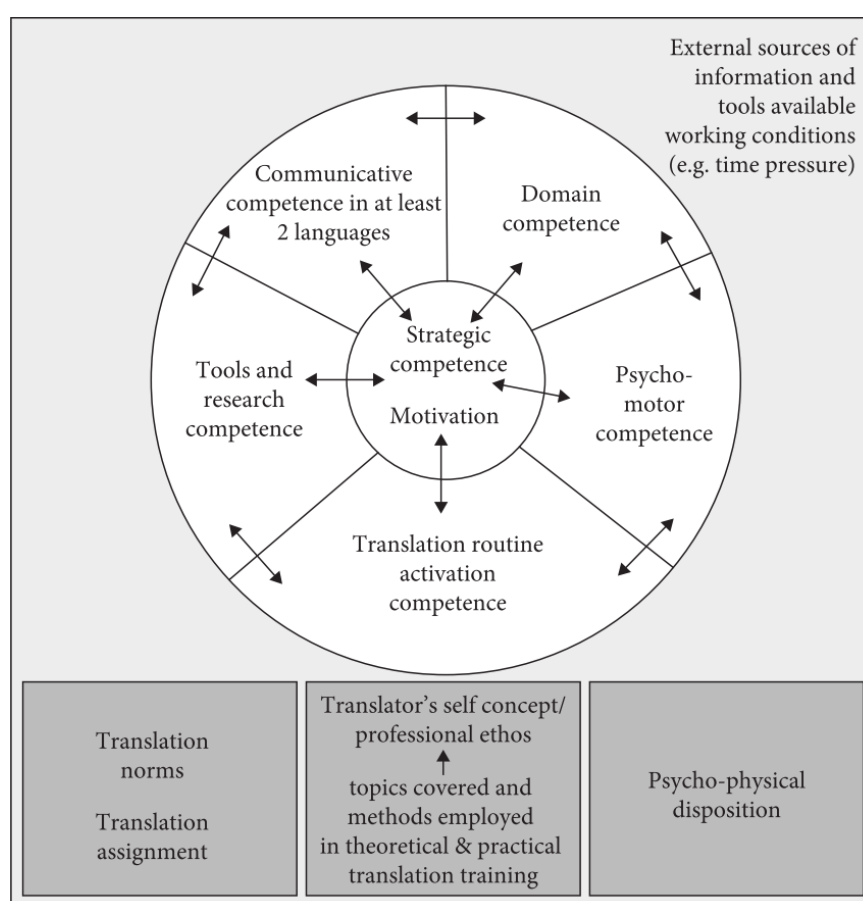
FIGURE 2.1: Translation Competence model (Göpferich 2008, p. 155, reprinted with the permission of the author)

Göpferich defines the six principal competences as follows:

1. *Communicative competence in at least two languages*
It comprises lexical, grammatical and pragmatic knowledge in both languages. Pragmatic knowledge also includes knowledge about genre and situation specific conventions in the respective cultures [...]
2. *Domain competence*
This [...] comprises [...] general and domain-specific knowledge. [...]
3. *Tools and research competence*
This [...] comprises the ability to use translation-specific conventional and electronic tools, from reference works such as dictionaries and encyclopaedias (either printed or electronic), term banks and other databases, parallel texts, the use of search engines and corpora to the use of word processors, terminology and translation management systems as well as machine translation systems.
4. Translation routine activation competence
This competence comprises the knowledge and the abilities to recall and apply certain - mostly language-pair-specific - (standard) transfer operations (or shifts) which frequently lead to acceptable target-language equivalents.

[...]
5. Psychomotor competence
These are the psychomotor abilities required for reading and writing (with electronic tools). The more developed these competences are, the less cognitive capacity is required, leaving more capacity for other cognitive tasks.
[...] 6.*Strategic competence*
This [...] controls the employment of the sub-competences mentioned above. As a meta-cognitive competence it sets priorities and defines hierarchies between the individual sub-competences [...].
(Göpferich 2009b, with competences in italics that directly correspond to sub-competences in the PACTE model)

Göpferich investigates the development of translation competence "in its continuity", based on Dynamic Systems Theory (DST; Thelen and Smith 1994, van Gelder 1998). She describes TC as a dynamic system, in which sub-competences interact in a phased manner over time. The subjects of her study are translation students. Hence, it is the experimental set-up and the underlying theoretical model that are of interest here, rather than the results. Göpferich states that the TransComp study concentrates on 1) strategic competence (the most important competence), 2) translation routine activation competence and 3) tools and research competence, as these are considered not to be inherent in bilinguals. She uses think-aloud protocols (TAP), as first deployed extensively in translation studies by Krings (2001), in order to collect data in an unstructured manner, enabling the collection of data that may not have been captured by her theoretical model. Göpferich finds that the development of the TC in the participants is non-linear and seems to stagnate at times. She hypothesises that some sub-competences need to reach a certain threshold before others can be improved, a concept that is inherent to the DST. Further, she suggests that there are factors, such as cognitive shifts, that are not covered by her theoretical model and data collection.

2.2.1.3   TC and Post-Editing (PE)

Göpferich's study shows that despite elaborate data collection methods, a controlled environment and variables in a longitudinal study, there are aspects that cannot be captured with her methodology and her theoretical model. She suspects the variables that constitute the missing links to explaining the development of translation competence to be rooted in other disciplines, such as developmental psychology.

Undoubtedly, post-editing skills traditionally differed from translation skills (see for example O'Brien 2002), as, for example, a translator produces the target text referring to the source text, while the post-editor edits an existing translation with or without access to the source text. However, translation and post-editing processes do share certain aspects, e.g. the correct transfer of meaning of concepts and relations. While the step of an initial draft translation by the translator is mostly skipped in post-editing (unless the MT output is discarded), translation problems faced by translators and post-editors may be similar (PACTE 2009). The question of whether translators and lay translators identify and deal with translation problems similarly was investigated by the PACTE group, who suggest that these processes are indeed similar (cf. section 2.2.1.1). While it has rightly been argued that traditional translation differs cognitively from post-editing (e.g. O'Brien 2002), with the advent of CAT tools and MT, the translation process has started to experience a cognitive shift, too. Especially with the prevalence of TMs and fuzzy matches, which also require editing, the cognitive gap between post-editing and translation has started to narrow. The TC model described above accommodates this shift only rudimentarily in the definition of the tools and research competence, which could be expanded easily. If it were to be applied to post-editing competence, deficiencies in language skills and knowledge of translation would

theoretically be compensated for by the machine translated output and post-editor domain knowledge.

For the purpose of this thesis, Göpferich's model is employed, which is largely based on, but constitutes a refined version of, the model by the PACTE Group. This thesis sets out to test the flexibility and applicability of this translation model to a *post-editing* scenario by lay translators in an online community setting. A potential basic model for post-editing competence would be useful in order to establish the required sub-competences for successful post-editors in general and for lay post-editing in the context of this thesis. With the resources available, the following competences of the Göpferich model will be investigated here: #1 communicative competence in at least two languages, #2 domain competence and #5 psycho-motor competence. Domain competence has been disregarded thus far in Göpferich's study but has been related to successful (monolingual) post-editing (Schwartz 2014), for example. We hypothesise that it will be of great importance in the context of community post-editing. Domain competence has been found to be a beneficial factor in monolingual PE recently (Aranberri et al. 2014, Schwartz 2014), as discussed in section 2.4.2.

The translation routine activation competence and strategic competence require research methods suited to their cognitive nature, e.g. TAP, which is beyond the scope of this thesis. Furthermore, research competence cannot be investigated, as only actions within the ACCEPT editor are recorded. The use of external tools was not encouraged as most participants did not have translation experience. Nevertheless, this thesis takes a step towards developing and testing a post-editing competence model and understanding why lay post-editing is successful/unsuccessful.

2.3   Improving MT

A need for a theoretical model of post-editing competence lies in the quality of current machine translation engines, which often makes post-editing a necessity. Despite advances in machine translation quality over the years, machine translated output has often been found not to be of publishable quality generally or of good enough quality for communication in online community environments specifically (Climent et al. 2003, Yamashita and Ishida 2006). More recently, there have been indicators that machine translated texts may be understood in online communities, e.g. with online community members rating 61% of MT texts as comprehensible on a binary scale (English to German) (Mitchell and Roturier 2012). However, there is still room for improvement, especially in MT for challenging language pairs, such as English into German MT, which struggles with differences in syntax. Several approaches have been taken and interventions have been made at different stages of the overall translation process in order to achieve translations of superior quality.

Firstly, a step towards improving MT output directly by changing the source text using pre-processing techniques, such as controlled language or pre-editing, has been investigated. This has been performed both in a manual (e.g. Roturier 2006, Doherty and O'Brien 2012, Bouillon et al. 2014) and/or automatic manner (e.g. Seretan, Bouillon and Gerlach 2014), most often identifying language-specific constructions and text characteristics that MT engines fail to translate accurately (e.g. Bernth and Gdaniec 2002, Monasterio 2009, Babych, Hartley and Sharoff 2009) and developing appropriate solutions to these translation problems. In addition to this, the portability and effectiveness of pre-processing solutions have been investigated (e.g. O'Brien and Roturier

2007).

The second and most obvious approach to improving machine translation output is to tune and enhance the MT engine used. While every translation setting involves different requirements, a widely used approach to building statistical machine translation systems in MT research at this time is a Moses-based (Koehn et al. 2007) engine that is optimised to the target domain. Examples of research strands include incremental retraining (e.g. Hardt and Elming 2010) and online adaptation (e.g. Denkowski et al. 2014). Furthermore, domain tuning and data selection methods have been the focus of work by different research centres working with machine translation worldwide. These are explored extensively in the nine workshops (2006-2014) on statistical machine translation that have been held in the field of computational linguistics, most recently, NAACL-HLT 2012, NAACL-HLT 2013 and ACL-WMT 2014.

2.4 Post-editing

A third approach to improving machine translated output occurs after the translation step and deals with correcting the machine translated output, i.e. post-editing. It is important to distinguish between two principal types of post-editing, automated post-editing (e.g. Knight and Chander 1994) and human post-editing. The chosen intervention here, which is facilitated by the ACCEPT portal (Roturier, Mitchell and Silva 2013) is community post-editing, a type of human PE. It constitutes a novel approach, as lay translators serve as post-editors, rather than professional translators. The ACCEPT project strives to provide a dynamic post-editing platform to facilitate quick turn-around times for translations, in order to optimise the user experience of community members. According to the commonly-used definition, post-editing can

be described as "checking, proof-reading and revising translations carried out by any kind of translation automaton" (Gouadec 2007, p. 25), which will be used for this thesis. Two types of PE are distinguished: rapid PE and full PE. Rapid PE (also minimal PE, gist PE or light PE) produces an output that is not necessarily of very high "publication" quality. It involves the correction of critical errors only; the focus is on speed. In contrast, full PE or conventional PE, involves the correction of minor and major errors with the aim of creating a higher quality product (e.g. Allen 2003, pp. 304–306), and usually one that is equal to the quality expected from human translators. In order to understand the requirements for PE interfaces and characteristics of PE in general, post-editing has to be clearly distinguished from the processes of translation. Evidence of practical differences between translation and post-editing is presented by Krings (2001) who demonstrates that post-editing differs from translation from a cognitive point of view. According to O'Brien (2002), the cognitive differences are based on the fact that conventional translation (without using CAT tools) draws on one source text only, while post-editing draws on two STs, the original and the raw MT output. As previously mentioned, there has been a shift triggered by the advent of TM and MT matches, which has assimilated the translation and post-editing processes in recent years. As pointed out, this has not yet been integrated into TC models. This shift should definitely influence and inform the forming of a revised translation and a new post-editing competence model. For the purpose of this thesis, however, this shift represented by TM and MT matches, remains embedded in Göpferich's tools and research competence. Going forward, it would be desirable to investigate the tools and research competence in particular, comparing translators and post-editors.

2.4.1   Bilingual PE

Comprehensive overviews on post-editing have been previously assembled (see O'Brien 2006a, de Almeida 2013); hence, a concise overview of themes and the latest research will be presented in the following, focussing on cognition, quality, productivity and PE tools.

2.4.1.1   PE and Cognition

Cognitive aspects, especially cognitive effort, have been studied since the early years of post-editing. According to Krings (2001), post-editing effort consists of a temporal, a technical and a cognitive component. While the first involves recording the time spent post-editing, the second involves recording the post-editors' keystrokes and mouse clicks and the third typically involves eye-tracking and/or TAP. Cognitive effort has gained particular attention for identifying whether PE can be a useful alternative or addition to conventional translation (see e.g. Guerberof 2012, Lacruz, Shreve and Angelone 2012, O'Brien 2011, Plitt and Masselot 2010).

O'Brien, for example, investigates the impact of NTIs (negative translatability indicators) on PE effort - temporal, technical and cognitive. In her experiment, nine participants post-edit two texts that have been machine translated from English into German, one that involves NTIs and one for which all NTIs have been removed. She confirms that NTIs increase the cognitive effort, as measured by CNA (Choice Network Analysis, i.e. different choices translators make while post-editing). O'Brien further investigates the relationship between pauses and cognitive effort (2006b). She reports that there are no significant differences between sentences that seem to be of better

machine translatability and those that seem to be less ideal. She concludes that pauses are very much subject to individual translator profiles - which is an observation that is made repeatedly in translation and post-editing research.

Tatsumi (2009 and 2010) provides an early attempt at predicting post-editing effort in terms of time with the help of AEMs (Automatic Evaluation Metrics) and sentence-complexity features. She finds, for example, that an ST complexity score based on the Systran[2] MT system correlates well with post-editing time.

Furthermore, Lacruz, Denkowski and Lavie (2014) investigate pauses as an indicator of PE effort in order to identify a measure for predicting cognitive demand by MT output, rather than cognitive effort during post-editing by dealing with edited-error to word metrics. They identify the transfer-error to word ratio as a better predictor than the mechanical-error to word ratio.

Koponen (2012) compares perceived post-editing effort to post-editing operations. She aims at teasing out methods of adequately measuring PE effort, by focussing on post-editing times on a sub-sentence level by five translators in the CASMACAT Work-bench.[3] Furthermore, she seeks to identify patterns in the source text that heighten cognitive effort, measuring duration, pauses, insertions and deletions for each sub-sentence Production Unit (PU), "a sequence of successive (insertion and deletion) keystrokes that produce a coherent passage of text" (Koponen 2012, p. 176).

Koponen finds that analysing patterns at sub-sentence level is very effective for commenting on PE effort and locating errors that are costly to fix. She shows, for example, that the existence of verbs in PUs was an indicator for higher cognitive effort in her

---

[2] http://www.systransoft.com/translation-support/v6-desktop-products/
[3] http://www.casmacat.eu/

experiment. Furthermore, she finds that sentence length is generally correlated to post-editing time. She acknowledges that her results are not clear-cut or generalisable. Her experiment shows that results may at times be contradictory and depend on translator individuality. It can be concluded that research in PE needs to be tailored to the environment in which it is produced and is rarely generalisable but may be transferable to similar experimental set-ups.

In order to address shortcomings, such as translator individuality, Vieira (2014) looks at both ST and MT output features and the post-editor profile (Working Memory Capacity (WMC) and SL proficiency). In this eye-tracking study, a pool of 13 translation students and professional translators post-edit two newspaper articles using the post-editing tool PET (Aziz, de Sousa and Specia 2012). On the way to identifying predictors for cognitive effort in PE, he finds that one MT automatic evaluation metric (AEM), Meteor, correlates with certain measures of cognitive effort. Furthermore, Vieira (2014) finds that a low SL proficiency results in a higher perceived post-editing effort in the participants. He also reports a connection between participants with low SL proficiency, who refer to the ST less frequently, and lower perceived cognitive effort. Vieira (2014) identifies a tendency that a higher WMC relates to increased productivity - pointing out, however, that there is a lot of variety amongst participants, which needs to be confirmed by further research.

The body of research presented on cognition in post-editing demonstrates that there is no standard way of measuring and analysing cognitive effort reliably, yielding consistent results. Nevertheless, it does demonstrate effective ways of predicting cognitive effort for post-editing.

2.4.1.2 PE and Quality

The concept of quality is complex and should be defined according to the context of the translation, as will be discussed in more detail in Chapter 3 section 3.2.2.5. In Translation Studies, quality is measured, for example, with the help of quality evaluation error typologies (de Almeida 2013), usability tests with potential end-users (e.g. Doherty and O'Brien 2013) and comprehension tests (as seen for translated quality in Sinaiko and Brislin 1973). Translators and their post-editing performance, e.g. the time they spent post-editing and the quality of the post-edited content (e.g. O'Brien 2006a, Guerberof 2009, Tatsumi 2010, de Almeida and O'Brien 2010) continue to be a key topic in post-editing studies.

Furthermore, studies compare quality of post-edited content to conventionally translated content and content translated with the assistance of CAT tools. Fiederer and O'Brien (2009) investigate how the quality of HT compares to post-edited MT content. They find that the post-edited content was assessed to be clearer and more accurate than the human translations, while the latter were evaluated to be of better style. Guerberof (2009) reports that segments provided by the machine translation engine employed and subsequently post-edited yield a higher quality PE output, as classified by the LISA QA model (Ray 2009), than those with TM fuzzy matches as a basis for translation. Garcia compares the quality and productivity of post-editing MT content to conventional translation (2010, 2011). He finds that there are only marginal differences in terms of productivity between the two, while the quality is clearly better for the post-edited output. In his later study, these results hold true regardless of the language direction (English-Chinese), text difficulty, or the translator's absolute performance (Garcia 2011).

O'Curran (2014) also compares post-editing quality of MT to post-edited content and reports that the errors annotated using a QA model based on the LISA QA model and SAE J2450 (SAE-International 2014) are similar across both translation types. Additionally, she reports that the human translations contain more errors in general and more stylistic errors.

In order to accommodate the need to consider a translation and its quality within its context, de Almeida (2013) focusses on translator individuality, by exploring the relationship between translator experience and post-editing success. She takes a step to uncovering similarities and differences in post-editing behaviour with twenty professional translators post-editing 74 segments in either French or Brazilian Portuguese from the IT domain. She finds that there are no correlations between translation experience and quality or post-editing experience and quality. She identifies other aspects, however, such as self-reported opinions on MT, that appear to be predictors for the post-editing quality produced by the translators, without reporting statistical significance in this matter. This demonstrates, again, that post-editing success is subject to the translator/post-editor profile and psychological aspects that need to be measured, in order to yield a complete picture of quality generation.

Returning to translator variance and the importance of psychological aspects in post-editing, Teixeira (2014) explores perceived translation quality with measured translation quality of ten translators (English to Spanish) without translation suggestions, with translation suggestions (TM and MT) and with metadata (e.g. source of the segment, fuzzy match level, and MT and TM suggestions). Using a mixed-method approach, he measures the translation time, edits and errors and conducts semi-structured interviews. He finds that the translators prefer an editing environment with integrated

suggestions and metadata to conventional human translation - a preference that he reports to be without correlation to quality.

These studies prove that post-editing may produce higher quality output, often assessed by error annotation, than human translation (Garcia 2010, 2011) and/or translation assisted with TM fuzzy matches (Fiederer and O'Brien 2009). They also demonstrate that quality may vary according to some aspects of the translator profile, e.g. attitudes and skills, but not others, e.g. post-editing experience, and expose the difficulty of accurately identifying translator or lay translator skills to perform post-editing successfully.

2.4.1.3 PE and Productivity

Post-editing quality and productivity are often investigated in combination, as demonstrated. Additionally, Plitt and Masselot (2010) compare productivity of human translation to that of post-editing of machine translated output in an industrial setting and report a significant increase in productivity. Similarly, Zhechev (2012) reports productivity gains varying depending on the language pair. While productivity gains are not always reported as being significant, the majority of studies suggest that MT output is helpful for increasing the translation throughput.

Most recently, Guerberof (2014) investigates translator productivity in editing both MT and TM matches. She finds that there is no significant difference in productivity and quality between editing MT or TM matches but that the productivity and quality were higher with the assistance of matches than for HT. It needs to be noted that the MT engine in her study produced raw MT output of good quality, which is not

surprising given the domain (IT) and language pair (English-Spanish) but which may not be realistic when dealing with the translation of UGC from English into German.

Aranberri et al. (2014) acknowledge that PE has mostly been researched with professional translators/translation students in mind, and seek to explore post-editing as an option for lay translators. They investigate post-editing productivity of six professional translators and that of six lay translators (university staff). Aranberri et al. find that there is a productivity gain, measured in time, across both groups, depending on the text edited. They report that the two groups of translators perceive MT output differently, with lay translators perceiving MT as more helpful than professional translators and that lay translators seem to benefit more from the machine translated output. They draw further connections between aspects of the post-editor profile, such as domain knowledge, where lay translators had an advantage over the professional translators, and the context, e.g. ST complexity and the quality of the MT output, which they find to be inter-connected and to affect the productivity gain without reporting statistical significance. This shows that productivity, similarly to quality, depends also on the translator profile.

In their study, Gaspari et al. (2014) point out the difficulties that arise with translator variance. They explore perceived versus actual post-editing effort and productivity gains. For this purpose, twenty translators working in the media sector post-edited texts taken from two bilingual wiki sites that had been translated with CoSyne (SMT engine; Martzoukos and Monz 2010). They find that perceived effort with MT seems to be at odds with the productivity measured for two out of the four translation directions. The discrepancy between perceived effort and real effort in regards to post-editing has prevailed in PE experiments, which may be due to professional translators' aversion to

machine translation and post-editing tools, as explored by de Almeida (2013).

These studies show that post-editing productivity may be higher, the same or lower than translation productivity and that it does not necessarily correlate with translator/post-editor sentiment. This gives some indication that post-editing productivity is highly dependent on the individual post-editor and that it may be influenced by factors that have not been taken into account in the experiments described above that are hard to measure because they are intangible.

2.4.1.4 PE tools

There have been efforts from a technical perspective to optimise the post-editing process and to minimise translator frustration. Intuitive and supporting (web-based) platforms for post-editing and post-editing research have been developed (e.g. iOmegaT - Welocalize and CNGL 2014, the ACCEPT Portal - Roturier, Mitchell and Silva 2013, CasMaCat - CasMaCat Consortium 2013, MateCat - MateCat Consortium 2013, TransCenter - Denkowski and Lavie 2012, Caitra - Koehn 2009) as well as standalone solutions (e.g. PET - Aziz, de Sousa and Specia 2012) (for a comparison of available tools see Vieira and Specia (2011)). This has also been expanded to interactive (e.g. Sanchis-Trilles et al. 2014) and intelligent post-editing features (e.g. O'Brien et al. 2014, Moorkens and O'Brien 2013) and mobile PE interfaces (Moorkens, O'Brien and Vreeke 2014).

Most recently, Sanchis-Trilles et al. (2014) test new functionalities, such as Interactive Translation Prediction (ITP), a dynamic text prediction functionality, and interactive translation prediction with advanced features (AITP) against conventional post-editing in terms of productivity using the Casmacat Workbench. In their analysis, they focus

on post-editing times, edit operations and gaze behaviour. While they find that using ITP requires fewer keystrokes but slightly more time than conventional PE, they find that the advanced features of AITP decreased productivity.

While some studies venture into investigating PE in the light of lay translators, none of the studies refer to a theoretical model of post-editing or translation competence, which is one challenge this thesis is seeking to address by taking a first step into testing whether the TC model described in section 2.2.1.2 may be applicable to lay post-editing.

With the availability of these platforms, post-editing has fought its way into the professional world of translation, for example through the option of displaying MT segments in CAT tools. Although post-editing still has mixed impacts on translator morale, it is undoubtedly here to stay. There have been more studies recently on investigating PE behaviour and sentiment (e.g. ACCEPT 2013a, Koponen 2013, Moorkens and O'Brien 2013, ACCEPT 2012) with a more positive perception of post-editing tools. Post-editing may not be accepted by all or even the majority of the translation world, but it has been observed that translators or translation students rarely reject machine translation results if they are presented with them (see for example Pym 2012). There have been tentative steps towards integrating post-editing into the translator training curriculum (O'Brien 2002, Belam 2003 and Depraetere 2010) within the field of Translation Studies. Steps have further been taken to transform and educate the new (and indeed the existing) generation of translators on post-editing and how to engage with it; for this purpose, post-editing courses (e.g. TAUS 2014) are offered, for example. The development of PE tools demonstrates the effort to optimise the post-editing process and to support all users and in all forms of translation.

As has been demonstrated, bilingual post-editing has been investigated from a number of angles, with research in advanced stages in topics, such as predicting post-editing effort. Monolingual post-editing, i.e. post-editing without reference to the source text, is addressed in the following sections.

2.4.2 Monolingual PE

Bilingual post-editing may not always be feasible, for example, if there is a lack of bilingual specialists or if the volume of content to be post-edited professionally in a bilingual manner exceeds the budget allocated by the company. In such scenarios, monolingual post-editing has been proposed as an alternative. Monolingual post-editing, the correction of machine translated text without reference to the source text, has been deemed impossible, due to the high cognitive effort involved (Krings 2001, p. 555). Further, the notion that a translation may be too inaccurate, if the MT system distorts the sense and if the post-editor is subsequently unable to extract the intended meaning and adapt the text accordingly (e.g. Koponen 2011), poses the question of whether monolingual post-editing can ever be successful. Nevertheless, monolingual post-editing has the potential of being more efficient and of opening PE activity to people who do not speak the source language, which could be a crucial factor in increasing the number of post-editors in an online community with a primary purpose other than that of translation.

Research in the area of monolingual post-editing has gained little attention, producing only few studies to date. Hu, Bederson and Resnik (2010 and 2011) develop MonoTrans, a system designed to facilitate monolingual translation by connecting a

monolingual translator of the source language with one of the target language via machine translation. It enables its users to detect and identify translation errors, which can then be resolved by adding additional information or by using back-translations. The process of machine translation, correcting, annotating and backtranslation is repeated until both the TL and the SL speaker are satisfied with the sentence. Hu, Bederson and Resnik (2010) conduct a small-scale project with six participants (lay translators). They find MonoTrans successful in helping the TL speakers to understand the sentences, even if the original MT output was of a very low quality. Furthermore, they report that in terms of adequacy, the number of high scoring sentences increases compared to raw MT output.

Based on the need for a more practical solution, Hu et al. (2011) develop MonoTrans2, which facilitates synchronous collaboration between two groups of monolingual translators. Connected via MT, multiple users can edit multiple sentences on both sides at the same time, eliminating any waiting time and making the process more efficient compared to the workflow in MonoTrans (Hu et al. 2011, p. 1). They rate two aspects, fluency and adequacy, on a 5-point scale and find that there is a higher throughput than with MonoTrans and that there is a significant increase in adequacy and fluency in comparison to raw MT output. While the experimental design, e.g. imbalance in translators per language pair, could have benefited from more rigour, the results clearly indicate that MonoTrans facilitates monolingual translation successfully, when compared to the raw MT output.

Koehn (2010) conducts a study comparing bilingual human translations and the output of monolingual post-editors to the output of monolingual post-editors using an

interactive machine translation tool. This tool provides translation options and suggestions on how to complete a sentence based on a phrase-based MT system, based on Moses (Koehn 2010). Without access to bilingual evaluators, the outputs were compared to reference translations according to a binary metric for fluency and adequacy combined. Koehn reports that one third of all sentences generated by post-editing and interactive machine translation with ten monolingual translation students as participants were rated as correct. Koehn states that the translation/post-editing performance varies across participants and texts. He concludes that general language skills and the amount of effort invested in the tasks are determining factors for the quality of the output. Drawbacks of this experiment lie in the small number of participants and monolingual evaluation based on a binary metric, which may not be adequate to catch slight differences in translation quality. It is noteworthy, however, that by using Koehn's approach, it is possible to create an acceptable translation without the participation of a bilingual translator.

Lin et al. (2010) present the idea of a Language Grid, a system that allows for a dynamic combination of 60 language resources in more than 50 different languages from various sources, including the internet, universities and research labs. They set out to determine whether integrating human tasks into the Language Grid, such as revising machine translation output or checking revision results, has a positive effect on the localisation process and cost. They find that translation quality can be successfully improved and that cost can be reduced. Again, this is a small-scale study with one translator and one monolingual reviser. Results are not of statistical significance, but provide insight into general processes, such as the interaction between human activities and translation services.

Research in monolingual PE has been and continues to be sparse. Schwartz (2014) conducts a study testing the hypothesis of whether domain knowledge increases the quality for monolingual post-editing. He conducts a study with one monolingual post-editor editing one scientific article. Schwartz finds that 87% of the segments were edited successfully, as per an adequacy measure assigned by one human evaluator. While the result is encouraging for monolingual post-editing, it has to be considered that it was a study with one post-editor. Nevertheless, the experience with domain knowledge as a decisive factor for monolingual PE quality is of interest to this thesis involving community members editing content from a familiar domain.

In summary, these small scale studies demonstrate that monolingual post-editing holds potential when additional aids, such as translation options are added and/or when domain experts are involved. However, the studies draw on source texts that have been written professionally. It remains to be established whether monolingual post-editing is feasible and effective on a larger scale, in real-life use cases for UGC.

2.4.3 Crowdsourcing Translations

With the emergence of Web 2.0, new forms of translation, e.g. crowdsourcing translations in volunteer communities have become a popular alternative to more conventional forms of translation. This trend coincides with the notion that users of the internet are not only passive consumers but strive to be contributors. While there are platforms which facilitate crowdsourcing of any kind of task involving financial compensation, such as Amazon Mechanical Turk (Amazon.com 2014) or Crowdflower (CrowdFlower 2014), the focus here is on volunteer crowdsourced translation. It is

facilitated through platforms for crowdsourced translations (see Mesipuu 2012), fan-subbing, volunteer translations both for non-profit and for-profit organisations (see e.g. Mesipuu 2012, McDonough Dolmaya 2012, Desiléts and van der Meer 2011, O'Hagan 2011, Baer 2010, Perrino 2009). The underlying concept is that of crowdsourcing, a term first coined by Howe, as an "act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call" (2006 [Online]). In the field of translation studies, a multitude of terms have been used to describe these translation scenarios, sometimes used interchangeably, including "collaborative translation", "crowdsourced translation", "community translation" (see O'Hagan 2011 for an overview). According to O'Hagan they mostly involve volunteers who translate content collaboratively, while being connected via a specific translation platform. Crowdsourced translation is commonly used to describe these processes. This concept has been transferred subsequently also to post-editing, i.e. crowd post-editing (see for example Tatsumi et al. 2012).

For the purpose of this thesis, we would argue that post-editing performed in the German Norton Community is not crowdsourced post-editing but rather community post-editing, for the following reasons. Crowdsourcing has been connected to companies translating their websites/products, such as Facebook, Twitter or Skype by building specific communities around the purpose of translation. The German Norton Community had already been established prior to the start of this project. The principal purpose of the community remains the discussion of the Norton products with

the by-product of post-editing by (mostly) lay translators. Pym's definition of "community translation"[4] emphasises the use of lay translators: community translation is a "(t)erm used for the practice whereby non-professionals translate software or websites that they actually use" (Pym 2011, p. 78). Community post-editing involves the community members, rather than professionals, post-editing software or websites and user-generated content that they actually use. Obviously, it cannot be ruled out that professional translators partake in either community translation or community post-editing. With the emergence of community PE, the pool of post-editors has been expanded to lay translators. These may possess varying levels of language skills, distinguishing it from crowdsourcing translation/post-editing for which quality control and user profiling is often performed (see Muntes and Paladini 2012, Yan et al. 2014, Unbabel - Graça 2014). This differs from the approach taken here, where we anticipate that community members will naturally address quality assurance, due to the self-governing nature of the community.

Using lay translators to do voluntary translation work has largely been subject to criticism from an ethical perspective. A common opinion is that voiced by Dodd, for example, "Crowdsourcing – the exploitation of Internet-based social networks to aggregate mass quantities of unpaid labor[5] – is just one more method corporations have found to push their costs onto the public while retaining profits and property rights for themselves" (2011). The topic of ethics in volunteer translation has been examined from the context of both professional and lay translators. Drugan (2011), for example, builds on McDonough Dolmaya's work (2011) and analyses different professional and

---

[4]It is of importance to distinguish 'community translation', which refers to translation in an online community from the concept of community translation that is also known as 'public service translation'. Thus, community translation throughout this thesis always refers to online community translation.

[5]As mentioned above, there are popular examples of paid crowdsourced translations, which is not the primary focus here.

non-professional codes of translation ethics in order to identify conflicts, contradictions and flaws within and across the same and gives an overview of topics that need to be addressed in the future with the continued rise of community translation.

From a practical perspective, crowdsourcing and crowdsourced translations have been met with mixed emotions by professional translators. Meanwhile, the business world is quick to assure that "companies are not implementing crowdsourcing to reduce their costs" (Ray 2009) but that they use crowdsourcing to involve the community and to provide translations for content that traditionally would not get translated. Crowdsourcing, albeit becoming more prevalent in the translation world, remains under scrutiny, which will undoubtedly continue to be the case.

The concept of ethics needs to be discussed briefly in the context of this thesis. The translations produced here are sourced from within the community for that same community. These post-editors have the advantage of being part of the community already; they are familiar with the field and the practices of the community. The topic of this thesis constitutes an attempt at bridging the gap between the language-specific community silos of the Norton Communities, thereby providing a richer user experience with a more extensive knowledge base, greater exposure of content and faster response rates to technological problems. In line with the arguments presented in Ray (2009), content that traditionally would not be translated is translated and can be seen by a broader audience with the potential of including additional languages and can help users who would not have access to solutions in their own language otherwise.

2.4.4   Towards Community PE

Crowd post-editing or "CPE", according to Tatsumi et al. (2012), encompasses the process of obtaining post-editing results through crowdsourcing. They indicate that the emphasis in research has been on facilitating crowd PE and developing frameworks for quality control, rather than examining the post-editing output itself. Crowd post-editing then, similar to crowd translation, happens in communities that have been specifically built for the purpose of post-editing and do not always involve volunteers but post-editors compensated for their work. Muntes and Paladini (2012) support CPE in an industrial context stating that it increases speed, resources and flexibility and reduces middle-man cost. They also describe quality assurance as the biggest challenge, which they address with a system of quality ranking and several steps of verification. Workers can be selected and closely monitored. Tatsumi et al. (2012, p. 2) also point out that CPE is more efficient in terms of cost and speed compared to HT; that it yields better results due to the members' domain knowledge; that it facilitates community development, but also mention quality assurance as a major challenge. They conduct a study in which they translate the university website of Toyohashi University of Technology into nine different languages using the Microsoft Translator's Widget and 21 students, native speakers of the target languages, to post-edit the output. The participants were paid for their efforts and did not have professional translation experience or linguistic expertise. They were not only able to edit the MT output but also the edited content of other students, i.e. they used a collaborative approach. Their evaluation involved picking the best translation candidate for a segment. Tatsumi et al. find that up to 10% (depending on the language pair) of the raw MT segments and up to 68% of the candidates that were edited last in the collaborative manner described

were picked as the best translation. They report that for up to 25% of the results, however, the evaluators were not satisfied with any translation candidates. This shows that crowd post-editing in a collaborative environment was successful for a majority of the post-edited content. In terms of design, this study differs from the present experiment described above in that the community specifically comprised students for the purpose of post-editing, while the Norton Community already exists with the purpose of discussing the Norton products. In addition, the post-editors used by Tatsumi were reimbursed for their effort, while the post-editors in this study are volunteers.

The research in crowdsourcing translation and its commentary demonstrates that it has been a controversial topic for ethical reasons and quality-related reasons, while it also demonstrates the power and ability of the crowd.

2.5   Concluding Remarks

Part I of this Literature Review introduced two central components for this research. Firstly, it discussed (the need for) a theoretical model for post-editing competence. Secondly, it introduced post-editing as a developing area of research, consisting of bilingual and monolingual PE, which has experienced a shift not only in the complexity of the topics studied but also in the platforms employed and the people who are targeted as post-editors, from professional translators to crowds online to lay translators. The current thesis seeks to take a first step in establishing a post-editing competence model by testing the applicability of sub-competences of Göpferich's TC model in regards to community post-editing by lay translators. It further aims at providing more insight into community post-editing by lay translators, a field that is only just gaining traction.

Chapter 3

Literature Review - Part II

3.1 Introduction

Part II of this Literature Review, Chapter 3 focusses on the quality of machine translated and post-edited content and its evaluation. The context for this experiment is an Online Community (OC), which is why the second half of this chapter is dedicated to investigating OCs in the light of their development, characteristics and previous research outside and within the field of translation studies. Evaluation of MT and PE content is discussed, initially considering the challenges that occur when evaluating the quality of machine translation and the approaches that have been developed to overcome these challenges. While both automatic and human evaluation are discussed, the focus is on the latter. The development of and current best practices in human evaluation are presented subsequently. This chapter then investigates Online Communities, outlining their history and development in general and as platforms for MT and PE in particular.

3.2 MT Quality Evaluation (QE)

The quality evaluation of machine translated content has been subject to extensive debate and development since the advent of machine translation systems, driven by the necessity to assess and improve MT engines. Three major strands of MT evaluation are automatic evaluation, human evaluation and usability evaluation. The development of each type of evaluation and current trends in research are explored in the following sections.

3.2.1 Automatic Evaluation

Automatic evaluation of machine translated content was first developed in order to test, tune and improve machine translation systems. Automatic evaluation metrics (AEMs) were considered ideal, as they facilitated quick and frequent evaluations of large test sets without requiring input from human evaluators. BLEU (Bilingual Evaluation Understudy) (Papineni et al. 2002) was the first AEM developed and continues to be the most commonly used metric. It measures the closeness of MT output to human translations by comparing a candidate translation against one or more reference translations. BLEU uses modified precision to execute this comparison with the help of n-grams (see Papineni et al. 2002, 312f). Meanwhile additional automatic measures, such as Meteor (Lavie and Agarwal 2007) and TER (Snover et al. 2006) have been developed, each of which focus on a different aspect, seeking to rectify shortcomings of the others. The following paragraph taken from Banerjee and Lavie (2005) summarises the challenges that automatic evaluation is facing:

> In order to be both effective and useful, an automatic metric for MT eval-
> uation has to satisfy several basic criteria. The primary and most intuitive
> requirement is that the metric have very high correlation with quantified
> human notions of MT quality. Furthermore, a good metric should be as
> sensitive as possible to differences in MT quality between different systems,
> and between different versions of the same system. The metric should be
> consistent ..., reliable... and general ... Needless to say, satisfying all of the
> above criteria is extremely difficult, and all of the metrics that have been
> proposed so far fall short of adequately addressing most if not all of these
> requirements. Nevertheless, when appropriately quantified and converted
> into concrete test measures, such requirements can set an overall standard
> by which different MT evaluation metrics can be compared and evaluated.
> (Banerjee and Lavie 2005, p. 66)

It is necessary for an AEM to be highly sensitive and correlated to human judge-
ment, in order to pick up nuances of differences between MT systems' performances.
Although human judgement is not necessarily consistent, it yields informed quality
results. Hence, a correlation between AEMs and human evaluation is desirable.

There have been studies stating that automatic metrics correlate well with human
judgement scores at paragraph and text level (Snover et al. 2006, Banerjee and Lavie
2005, Coughlin 2003, Doddington 2002, Papineni et al. 2002) but not on a sentence
or segment level (Volk 2008, Lavie and Agarwal 2007, Callison-Burch, Osborne and
Koehn 2006, Snover et al. 2006, Turian, Shen and Melamed 2003). While progress
has been made in this area, a correlation in the circumstances of this experiment is by
no means guaranteed, which constitutes a deciding factor against AEMs as primary
quality evaluation measures in the full-scale study for this thesis.[1]

As presented above, AEMs may be useful in evaluating and tuning machine translation
systems. Starting out with BLEU, effort has been made with growing success to

---

[1]To confirm or dismiss the possibility of a correlation, Meteor and TER were employed in the pilot
study, which yielded results dismissing correlations between human evaluation and AEMs for the data
gathered. The result are presented in Mitchell, Roturier and O'Brien (2013). TER is used as an edit
distance measure between the MT and the PE output for the qualitative evaluation as presented in
Chapter 6, however.

correlate AEMs with human evaluation in order to evaluate MT output consistently and efficiently.

While one might argue the case for and against automatic evaluation, it was decided not to use AEMs as a primary method for quality evaluation in the full-scale study presented in this thesis. The reasoning behind this decision is as follows: the experiment in this thesis deals with a source language style that is very informal. Human reference translations might fail to reproduce this style (sentence length, word order etc.) in which case the post-edited content (even if not wrong) will differ greatly from the human translations resulting in a lower BLEU score. Furthermore, automatic metrics do not consider the circumstances of the translation, such as, the recipient's characteristics, expectations and demands.

Human evaluation may offer a higher degree of control over quality definition. Quality is not absolute; however, automatic metrics are based on predefined formulas to assess quality. We assume that volunteers untrained in the field of translation with varying language skills produce unpredictable output. We take a mixed-method approach to QE in this thesis, to facilitate an in-depth analysis in addition to a more high-level (quantitative) analysis of the post-editing quality produced (see Chapter 4). For the empirical analysis, human evaluation is the approach that will yield a comprehensive overview of the quality produced by volunteer post-editors in the Norton Forum. Human evaluation is considered of particular value here, as it involves the end-users of the translations in the evaluation process.

3.2.2   Human Evaluation

The ALPAC report in 1966 was an early example of research into human evaluation of MT. It described the status of MT and its anticipated potential use at the time. The report aimed at developing new evaluation methods to assess aspects such as efficiency, quality and cost. While this first attempt at predicting the future of MT produced a bleak verdict, it set out to develop an effective and meaningful set of evaluation characteristics on which to judge MT output, such as fidelity and intelligibility. Since then, human evaluation has been studied extensively and has been developed further. However, research to date has failed to produce an official standard for human evaluation of machine translated and post-edited output.

3.2.2.1   Adequacy and Fluency

Two concepts that emerged very early on (e.g. van Slype 1979) and that have been used consistently over the years are adequacy and fluency (Denkowski and Lavie 2010). These concepts and their respective scales have been clearly defined by the Linguistic Data Consortium (2005). Adequacy and fluency are measured on Likert scales by human subjects in the following way:

> How do you judge the fluency of this translation?
> It is:
> 5 Flawless English
> 4 Good English
> 3 Non-native English
> 2 Disfluent English
> 1 Incomprehensible
>
> How much of the meaning expressed in the gold-standard translation is also expressed in the target translation?
> 5 All
> 4 Most
> 3 Much

2 Little
1 None (LDC 2005, para. 7f)

The importance of these two aspects is also reflected in the development of AEMs. For example, Snover et al. developed an error measure, TER (Translation Error Rate) based on the Word Error Rate (WER) (2006). It encapsulates both adequacy and fluency, as it counts the number of edits performed by a human to change machine translated content so that it is fluent and expresses the correct meaning.

Adequacy and fluency were also used for the evaluation task in the Workshop on Statistical Machine Translation (WMT) shared tasks (e.g. Koehn and Monz 2006), which focussed on improving SMT in collaborative research teams for a number of languages. Callison-Burch et al., for example, state that greater emphasis was placed on human evaluation than on AEMs, as they are "an imperfect substitute for human assessment of translation quality" (Callison-Burch et al. 2009, p. 6). They found it to be impractical to use multiple concepts such as adequacy and fluency to tune automatic metrics, as they needed to be combined in some meaningful way and as human evaluators had difficulty in distinguishing between the two aspects (Denkowski and Lavie 2010, p. 3). Denkowski and Lavie report that for a "translation to fully express the meaning of a reference, it must also be fully, or near fully fluent, as slight changes in word order and morphology can dramatically alter meaning in many languages" (Denkowski and Lavie 2010, p. 3). The fact is, that both meaning and fluency are valuable in the assessment of machine translated/post-edited content. While it is true that human evaluators can introduce noise and have difficulty in distinguishing between the different criteria, the two should not necessarily be combined. On the contrary, it is important to separate them, to eliminate ambiguity of whether a translation candidate obtained a low

adequacy score due to the language (spelling, grammar, style issues) distorting the meaning or due to the fact that the source was mistranslated by the MT engine and thus the meaning is wrong.

To solve this problem of seeking to combine two different criteria, they employed relative ranking in the WMT shared tasks (see for example Callison-Burch et al. 2008) instead. While relative ranking is often used to compare several machine translation engines (e.g. Callison-Burch et al. 2009), the two criteria adequacy and fluency are still commonly used to assess machine translation and post-editing quality.

These will be employed here in order to gain in-depth insight into segments independent from each other, rather than relative to each other in comparison to a reference translation. The concept of adequacy is complex and is used interchangeably with accuracy, fidelity, correctness or precision (King, Popescu-Belis and Hovy 2003) at times. It is difficult to clearly distinguish between these closely related if not congruent concepts. This dilemma becomes apparent when studying definitions of the same, e.g. a definition of fidelity as the "M[m]easurement of the correctness of the information transferred from the source language to the target language" (van Slype 1979, p. 72), which contains another term mentioned above, i.e. correctness, rendering it part of fidelity rather than an autonomous concept. Based on this definition of fidelity, it is a concept based on precision (another concept mentioned above) rather than recall. For the purpose of this experiment, it is important that recall is prioritised, as content is may be dropped or added in during the machine translation process, which may then be amplified or rectified by the post-editor, depending on their skills. A definition that was deemed an appropriate solution was that of the LDC, i.e. "Adequacy refers to the degree to which information present in the original is also communicated in the

translation." LDC (2005, para. 6) and the subsequent scale presented earlier in this section, as it encompasses the amount of content transferred.

It needs to be noted, however, that at no point during the study was the English term 'adequacy' used. It has been used solely in its translated form "Vollständigkeit", which was chosen after careful research and consideration, to both mirror the concept of adequacy in regard to recall and to be an accessible term for evaluators untrained in linguistics, translation or translation evaluation. Furthermore, the concept of fluency was presented to the evaluators in its translated German form. For this purpose, both "sprachliche Qualität" and "Sprachfluss" for the domain specialist evaluation was used, as there was no one satisfactory translation of "fluency" in German. For simplicity, only the term "sprachliche Qualität" was used for the community evaluators.[2]

3.2.2.2   Usability evaluation

Usability evaluation takes a practical approach to evaluation MT quality, as for example described in earlier years by Taylor and White (1998). Currently, usability evaluation is performed by assessing how MT quality affects the performance of human subjects. This typically involves answering questions or performing tasks expressed in the translated content (e.g. Castilho et al. 2014, Parton et al. 2012) and the performance of automatic applications, such as in regards to information retrieval (e.g. Parton et al. 2012, Parton and McKeown 2010). Usability evaluation provides insight into translation quality from the perspective of the end-user, which is extremely valuable in determining whether a translation is adequate or not in the context for which it is produced. However, usability evaluation can be subjective, time-consuming and

---

[2]Acceptability of the term "sprachliche Qualität" had been tested with six German native speakers unfamiliar with the study.

costly, i.e. tasks have to be developed and the necessary equipment/conditions for executing the tasks need to be provided. While usability evaluation would have undoubtedly been of use to this thesis, it was deemed too costly to monitor the subjects, provide the required equipment and reproduce the errors outlined in the translated texts, especially since the texts were picked at random and the participants remote.

Based on the research in human evaluation, we can say that it has been used successfully - admittedly not without struggles in resources and subjectivity - for a variety of purposes, such as in the translation industry and in small-scale and large-scale machine translation research projects. A relatively new development in this area is that of crowd evaluation. This type of evaluation is of particular interest, as it leads towards the novel aspect of evaluation for this thesis.

3.2.2.3 Crowd Evaluation

Human evaluation, especially in the case of evaluators untrained in translation/linguistics has drawbacks, as it is always subjective and often leads to inconsistent ratings, expressed in low intra- but especially low inter-annotator agreement (e.g. Lommel, Popovic and Burchardt 2014, Callison-Burch et al. 2007). Kirchhoff, Capurro and Turner summarise the struggles of human evaluation as follows: "When assessing a complex 'product' like MT output, users are notoriously poor at analyzing their own judgments and stating them in explicit terms, especially when they lack linguistic training" (2014, p. 6). Furthermore, they claim that even (relative) ranking as commonly used, for example, in WMT shared tasks (cf. Denkowski and Lavie 2010) still does not solve the problems faced in human evaluation. Despite these drawbacks, human evaluation still provides valuable in-depth insight into quality, which AEMs are not

designed to yield. Recently, crowd evaluation has become more popular in the field of translation and machine translation and is often conducted with the help of Amazon Mechanical Turk (AMT). It was launched in 2005 to meet the need for a platform to crowd-source content that is based on tasks (Human Intelligence Tasks - HITs) that require human input and cannot be completed by machines. Crowdsourcing on AMT has faced criticism similar to crowdsourcing in general and crowdsourced translations, including the abuse of workers' rights and negative impacts on both the future of work and technology (Ray 2009). Despite this criticism, AMT has been successful in convincing people of its benefits and has expanded into a massive workforce with a worldwide presence.[3] AMT now plays a major role in both industry and research. Cushing pinpoints it, when she writes "In seven years, Turk and its imitators have gone from experiment to emerging field to major industry" (2013 [Online]).

Crowd evaluation needs to be distinguished from community evaluation. Crowdsourcing, e.g. in Amazon's Mechanical Turk, can involve a reimbursement of the evaluators, whereas community evaluation does not. Furthermore, the crowd evaluation takes place on a separate platform (e.g. AMT) with workers who do not know the domain and who have never visited the forum before. Community evaluation in this study involves the end-users of the translations delving into the perception of quality and post-edited content within the target community. Lastly, the motivation to rate content is not financial but is driven by other reasons, such as community loyalty, or strong feelings about machine translation/post-editing. With a community evaluation, valuable feedback can be gathered, but not without running the risk of low participation or domination of the results by internal community sub-cultures.

---

[3]Since 2012, new memberships are restricted to the USA.

3.2.2.4   Combining QE Approaches

Human evaluation of translated content has often been criticised for being subjective and unpredictable. A reason for difficulties to achieve agreement amongst human evaluators may lie in individual variance and characteristics and a discrepancy between emotions/thoughts and reality, as observed in previous studies (e.g. de Almeida 2013; Koponen et al. 2012). In order to overcome bias by only choosing a single approach to quality evaluation, researchers have opted to combine several approaches. An early study that already takes a comprehensive approach to translation evaluation is that of Sinaiko and Brislin (1973). Their article, similar to findings presented in Chapter 2 section 2.2.1.2, states that there are causes for translation competence that must be rooted in the psyche of a translator that cannot be measured with traditional methods.

Sinaiko and Brislin (1973) approach the subject of translation quality and human evaluation from the field of Applied Psychology. Twelve translators translate three samples of technical manuals from English into Vietnamese. To gain a comprehensive view of translation, they employ back translation with an error categorisation, knowledge testing and performance testing. On the back-translation, they use a basic adequacy error categorisation to assess the quality. For knowledge testing, they employ two phases of reading and answering comprehension questions. With regard to performance testing, six teams perform tasks using either the original English instructions or the Vietnamese translations, the results of which were subsequently assessed and compared. They identify performance testing as the best evaluation technique and unfortunately the most expensive and time-consuming one.

Sinaiko and Brislin report close agreement for the error annotation results between

the two evaluators and same order in ranking for both knowledge and performance testing. They find that none of the translations, including the professional reference translations, achieved full marks for any of the three evaluation methods. The average quality score that the best translations received across all three methods was 73%, which serves as a benchmark for the (average) maximum quality achieved.

Sinaiko and Brislin also touch upon subjectivity in their study. They report that the Vietnamese airmen, who followed the best translated material, disliked the translations, although their performance was equivalent to that of the control group of native English speakers following the original instructions. Interviews revealed that there is a discrepancy between subjective assessment of translations and performance testing. This discrepancy between perceived effort or performance and actual performance has been reaffirmed in recent studies (Gaspari et al. 2014, Teixeira 2014). Of course, it has to be considered that performance may have been influenced by other aspects, e.g. experience, or that, for example, an unidiomatic translation may not distort the accuracy of the translation (cf. Doherty and O'Brien 2013, who find that even raw MT instructions were usable). It can be concluded that different types of evaluation may not always be correlated but they are valid in their own right, as it is not only important that the translations are correct but also that they are not irritating to the end-user. This needs to be considered especially when dealing with MT output, as its mistakes may seem minor to an evaluator but may be irritating to a user or indeed vice versa.

3.2.2.5 A Perfect Quality Evaluation?

Having presented different types of evaluation approaches and their advantages, disadvantages and some examples, the question of whether there is a perfect way of evaluating machine translated or post-edited content arises. The short answer to this is, no, there is not. Otherwise, a standard for evaluating translation quality would have already been established, accepted and widely used. There have been extensive efforts towards providing a customisable framework to assess translation quality both for the industry and for research purposes. Drugan's book on Quality in Professional Translation, for example, approaches quality in a comprehensive manner outlining numerous strategies of how translation companies and individual translators deal with measuring and improving translation quality. She divides these strategies broadly into top-down approaches, i.e. hierarchical models traditionally used in the translation industry and bottom-up approaches, more flexible approaches informed by the changes in requirements for a quality evaluation in order to deal with user-generated content, for example. While the first strategy has been criticised for being too rigid and therefore unsuitable for newer translation contexts, the second appears to still require some aspects of a top-down approach, e.g. a community manager. Nevertheless, Drugan argues that both top-down and bottom-up approaches continue to constitute important strategies, which constitutes one reason for employing both types for the purpose of this thesis.

Furthermore, attempts have been made to create standards to facilitate standardised quality evaluation of machine translated or indeed post-edited content for both the translation industry and researchers. Melby et al. for example, strive to create a formal

method of evaluating the performance of post-editors by proposing and testing formalized structured translation specifications (FSTS). They also recognise that "quality is not transcendent but relative" (Melby et al. 2012, p. 2) and base their specifications on five categories, source, target, production, environment, and relationships covering the circumstances and specifications of any translated or post-edited product. Their study demonstrates that lay translators can reliably evaluate post-edited content based on the framework developed. Despite various efforts of reconciling the complex environments in which translations happen into one framework, the debate on translation/post-edited quality continues. One reason for this is that translation quality is and will always be subjective and dependent on the circumstances, such as audience and the timespan the translation will be visible or useful for. With crowdsourced and community translation and post-editing comes a larger number of approaches that may be based on personal preference and on training or the lack thereof in relevant fields, such as linguistics. A solution to dealing with this variety and dependencies of quality evaluation is to offer several methods of quality evaluation, in order to create a comprehensive picture of the output to be examined. This has been picked up on by TAUS, who have been working on a Dynamic Quality Framework (DQF) in collaboration with 50 companies since 2011 (TAUS 2013a), in order to create a standard framework that is versatile in adapting to the users' requirements. With the help of the DQF tools and knowledge base, users of DQF can specify characteristics of the content to be evaluated, such as content type, e.g. marketing material or social media, or content channel, e.g. business-to-business, business-to-consumer etc. (TAUS 2013b). The DQF tool then suggests several quality evaluation models, including usability evaluation, error typology, adequacy/fluency and community-based evaluation. Similarly, as part of the QT Launchpad project (QT Launchpad Consortium 2014), which is striving to overcome

quality barriers in MT, HT and language technologies, an open source tool for customisable quality assessment has been developed, Multidimensional Quality Metrics (MQM) (see Lommel 2014 for an overview), in order to address the one-size-fits-all quality evaluation dilemma and to make it accessible to everyone.

As is evident, the evaluation of machine translated and post-edited content has been studied and tested extensively for rigour and appropriateness. With this thesis, we hope to extend this knowledge base by combining different evaluation techniques that have been used previously and applying them in a new context, that of post-editing of user-generated content in an online community by lay translators. Further, this thesis seeks to take evaluation one step further by exploring the involvement of end-users, members of the community largely untrained in translation, post-editing and evaluation, in the evaluation process.

3.3   Online Communities (OCs)

In order to understand PE competence and quality outside a traditional translation environment for lay translators, the platform that provides the lay post-editors, here an online community, needs to be investigated.

As noted in Chapter 2, the internet user's role has changed from being solely an information consumer to being an information provider. Today, blogs, wikis, podcasts and any type of social interaction serve the purpose of providing content (Iriberri and Leroy 2009). Often, these structures are built around online communities. The idea of an online community was first predicted by J.C.R Licklider and Robert Taylor in 1968, who launched the first OC. They anticipated the nature of online communities to be

the following: "In most fields they will consist of geographically separated members, sometimes grouped in small clusters, and sometimes working individually. They will be communities not of common location, but of common interest. . . ." (Rheingold 2000, p. 9).

According to Iriberri and Leroy (2009, 11:2), the WELL (Whole Earth 'Lectronic Link), established in 1985 was the pioneering OC. Rheingold (2000) describes this online community that provides services, such as discussion forums and private messages, as a source of information and also as a place to form social relationships.

3.3.1 Definition and Classification

Many OCs have been formed and evolved since the early days. The definition of OCs presented by Lee, Vogel and Limayem as a "cyberspace supported by computer-based information technology, centered upon communication and interaction of participants to generate member-driven contents" (2003, p. 51) is used for this thesis.

While member-driven content is the basis that all online communities share, communities differ in their purpose. Hagel and Armstrong (1997) define four categories of purpose: interest, relationship, fantasy and transaction. While more extensive classifications (e.g. Jones and Rafeli 2000) have been developed since, Hagel and Armstrong's classification is still widely used (cf. Lee, Vogel and Limayem 2003 for a basic categorisation) and is suitable for this thesis. It distinguishes clearly between the most widely known type of online community currently, i.e. a community of "relationship", with social network sites such as Facebook serving as an example, and the kind of community this thesis focusses on, the Norton Community, an example of a community of common "interest". Members are users of the Norton products and thus have a

common interest in Norton products, on which they share knowledge and opinions. Furthermore, it has been established that online communities evolve in stages following a life-cycle (e.g. Wenger, McDermott and Snyder 2002, Iriberri and Leroy 2009). Iriberri and Leroy base their system on a model capable of describing the development of any information system, called the information systems life-cycle (ISLC), consisting of the following stages: inception, creation, growth, maturity and death (Iriberri and Leroy 2009, 11:13). The original English-speaking Norton Community was launched in 2008 under the auspices of Symantec, followed by the German Norton Community later in the same year, both of which fall into the growth stage. They have passed through stage 1, inception, in which users start to interact with each other and stage 2, which involves creation, members have created a culture and identity of the community as roles have been allocated. They now reside in stage 3, growth, in which rules have established and subgroups are formed (for categorisation cf. Iriberri and Leroy 2009, 11:13f).

Just as with offline communities, online communities are characterised by complex hierarchical structures, which differ from community to community. The Norton Community is managed on a meta level by the community administrator and monitored by Symantec employees. It is, however, widely self-governed in that the super-users ("gurus", who enjoy additional rights and privileges in the community) report and deal with destructive or unhelpful behaviour in the community, with the administrator serving as an advisor to them and with the power to appoint new super users.

3.3.2 Research in OCs

The rapid development and growth of online communities affected multiple waves of research interest. Firstly, research was conducted in the field of sociology, exploring the potential of online communities to alter the communication and interaction of societies in the physical world (e.g. Wellman et al. 1996). Secondly, studies in management focussed on the member-driven content generated in OCs in order to establish its value to businesses (e.g. Hagel and Armstrong 1997). Thirdly, OCs invoked interest among researchers, especially psychologists. Their focus was on relationships and attachments between members in online communities (e.g. Blanchard 2004). Fourthly, research responded to a need to bring order to the relatively new and very dynamic field in order to understand and benefit from it (Reid and Gray 2007). Hence, information system research strove to create a framework to facilitate the implementation of studies (e.g. Lee, Vogel and Limayem 2003). In recent years, research in online communities has branched out further including research in social network analysis, for example, including sub fields such as data mining and social network modeling (Du et al. 2007).

3.3.3 Machine Translation in OCs

More recently, and especially with the advent of crowdsourcing (cf. Chapter 2), online communities have been used as platforms for investigating and improving machine translation. The following two studies investigate machine translation in combination with online communities (i.e. on a virtual campus – Climent et al. 2003) and in combination with tools used in OCs (i.e. a chat function – Yamashita and Ishida 2006). In their study, Yamashita and Ishida set out to establish how MT affects

communication online (2006, p. 515), using a system that could potentially be used in OCs. They use a multilingual chat system (Japanese/Chinese/Korean ↔ English) that instantly translates messages between two monolingual participants and focus on analysing referential communication, i.e. giving explanations and asking questions. They compare their findings for participants of different native languages using a chat system with and without MT. They find that in communication through a machine-translated chat system, even basic exchanges result in difficulties. Yamashita and Ishida further observe that participants attempted to compensate for disruptions by waiting for the speaker to offer more information, causing time delays, and by using the same vocabulary (if established that it is translated well) throughout the experiment. They conclude that MT as a channel for communication is not suitable for everyday use and suggest that an additional form of support is needed (Yamashita and Ishida 2006, p. 522). They point out that their study was prompted by a lack of research conducted in the area of MT as a dual-directional tool (Yamashita and Ishida 2006, p. 522), e.g. MT as an interactive tool in communication between humans. While using MT in a chat-like function has been researched since (e.g. Hu, Bederson and Resnik 2010, Hu et al. 2011), other forms of interactive MT, where the user interacts with the MT engine, for example by choosing a machine translated phrase from a list of options dynamically, rather than two users interacting through the engine, has gained in popularity, such as with the CasMaCat or MateCat systems (CasMaCat 2013; MateCat 2013). The ACCEPT project strives to connect (monolingual) speakers of different languages via an MT engine channel, which is ultimately facilitated with the aid of post-editing by native speakers. This thesis focuses on examining whether such post-editing is a feasible solution.[4]

---

[4]Feasibility in this thesis refers to post-editing producing output that is of superior quality than that of raw MT output.

Climent et al. (2003) conduct research in the area of MT in bilingual newsgroups in order to preserve Catalan in online communication, focussing on the machine translation of e-mails. They see the incentive for their research in the circumstance that existing MT systems solely work with correct and standardised input text, while, the genre of e-mails is characterised by noisy input, such as misspellings and creative language (Climent et al. 2003). Their research objectives include the desire to extract the features that pose difficulties for MT systems, and to analyse to what extent they are intentional. A macro-evaluation is carried out on general aspects of the MT system's performance, using criteria such as fidelity, readability etc. A micro-evaluation is conducted by human evaluators in order to point out the limitations of the MT system. They identify bad spelling and non-standard terminology as the two most challenging features for the MT system. They conclude that for both languages (Spanish and Catalan), the MT system requires customisation, such as adding further (non-standard) terms to the system, in order to be able to produce an acceptable output. Both studies conclude that the MT system is not sufficient in itself but that either additional support (Yamashita and Ishida 2006, p. 522) or customisation of the system, such as adding further terminology (Climent et al. 2003) is required. We hope to address these shortcomings in the experiment reported in this thesis by employing a light pre-editing step and post-editing as the principal human intervention facilitating better translation output.

In recent years, research has been conducted on post-editing of machine translated content by volunteers in online communities that have translation as their primary purpose (e.g. Rickard 2009). Further, volunteer translation has been studied in large

and very successful online communities. These communities had various primary purposes, such as social relationships, who formed sub-communities for translators with extensive quality control, the translation of Facebook (see Dombek 2014) being an example. While Dombek focusses on translation and a separate translation community within the Facebook community for which quality control is employed by community members and Facebook, this thesis focusses on post-editing by (potentially) a whole community with unknown skills.

3.4 Concluding Remarks

Following an introduction of the concept of post-editing itself, this chapter introduced the evaluation of machine translated and post-edited content demonstrating its complexity in design and challenges in producing a rigorous and appropriate evaluation approach. In order to preface novel aspects in this thesis, especially in the evaluation approach chosen, as outlined in Chapter 4, this chapter gave an overview of online communities as a new platform for research that has gained importance not only for translation but also for post-editing and the evaluation of MT and PE content.

Chapter 4

Methodology

4.1 Introduction

This chapter presents the research approach taken and the experience gathered from a pilot study and develops the methods deployed in the full-scale experiment. Section 4.2 introduces the concept of mixed methods research design and the reasoning behind this choice. Section 4.3 outlines the objectives, experimental design and results of a pilot study. The finalised methodology is subsequently presented in section 4.4. It outlines the research questions, followed by the independent/dependent variables arising from these and the hypotheses to be tested. The experimental design, procedure and methods to measure the variables of interest are discussed in sequential order. Subsequently, concepts are introduced that were addressed prior to the post-editing activity, for example pre-editing, user profile etc. Then, the experimental design and procedure of the post-editing process are described, including the user interface of the ACCEPT portal and the mechanism for collecting post-editing data and actions taken

by the participants in the portal. Finally, this chapter presents the methods employed to measure post-editing quality.

4.2 Mixed Methods Approach

The following section focusses on RQ1 and RQ2, as introduced in Chapter 1 and the concept of quality. RQ3 and RQ4 are not discussed here, as RQ3 is addressed in a purely quantitative manner and RQ4 on a theoretical level.[1]

As is evident from the diverse range of topics discussed in Chapter 2, community post-editing and the quality achieved by this activity is an interdisciplinary concept, which requires a flexible research design. A mixed methods approach is adopted in this experiment in order to gain a comprehensive picture of quality in an attempt to respond to that interdisciplinarity.

Historically, mixed methods research designs have developed from focussing on mixing methods only, while dissociating methods from philosophical worldviews (Greene, Caracelli and Graham 1989) to a more holistic form of mixing, involving methods, philosophical worldview and interpretations of the data (Tashakkori and Teddlie 1998). Efforts have been made towards a standard definition of mixed methods research. Creswell and Plano Clark, for example, define it as follows:

> As a methodology, it [a mixed methods research design] involves philosophical assumptions that guide the direction of the collection and analysis and the mixture of qualitative and quantitative approaches in many phases of the research process. As a method it focusses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of

---

[1]RQ1) How does post-edited output produced by community post-editors compare to raw machine translated output? RQ2) What are segment characteristics and post-editing patterns that render post-editing successful/unsuccessful? RQ3) Do elements of the post-editor profile serve as predictors for the quality of the post-edited output? RQ4) Is Göpferich's Translation Competence model applicable to community post-editing?

studies. Its central premise is that the use of quantitative and qualitative approaches, in combination, provides a better understanding of research problems than either approach alone. (Creswell and Plano Clark 2007, p. 5).

The idea that mixed methods research facilitates a better understanding than either quantitative or qualitative research alone has been widely debated (Creswell 2011). For the purpose of this study, we hypothesise this to be true. After a high-level quantitative analysis, it is beneficial, essential even, to investigate post-editing quality on a detailed level with focus on the actual edits made in order to assess post-editing quality from a linguistic perspective. In addition, mixed-methods research design approaches have been successfully established within the framework of Translation Studies (e.g. Moorkens 2012, Doherty 2012) in general and translation research in online communities in particular (e.g. Dombek 2014). In their newer work, Creswell and Plano Clark propose the following set of statements to be true of a mixed methods research design, which were used as a starting point for the present research design:

In mixed methods, the researcher

- collects and analyses persuasively and rigorously both qualitative and quantitative data (based on research questions);
- mixes (or integrates or links) the two forms of data concurrently by combining them (or merging them), sequentially by having one build on the other, or embedding one within the other;
- gives priority to one or to both forms of data (in terms of what the research emphasizes);
- uses these procedures in a single study or in multiple phases of a program of study;
- frames these procedures within philosophical worldviews and theoretical lenses; and
- combines the procedures into specific research designs that direct the plan for conducting the study. (Creswell and Plano Clark 2011, p. 271)

Based on the set of statements presented above, this study takes mixed methods research, as described in the following:

- It asks what level of quality community post-editors produce compared to raw machine translated output (RQ1) and *how* this quality is produced, i.e. which edits lead to high-scoring vs. low-scoring segments (RQ2). It therefore collects the post-edited (product-based) solutions and evaluation data (adequacy and fluency scores). While the latter are used to classify the level of quality achieved in a quantifiable manner, a sub-set based on the scores obtained is used to explain those empirical results.

- This research experiment takes an explanatory sequential design. Based on the quality scores obtained (quantitative data), the best scoring and worst scoring segments are identified from the set of post-edited segments (used for a qualitative analysis). These serve to explain the quantitative findings, i.e. the high and low quality scores.

- It gives priority to the quantitative data, as it is the method which collects more data and which renders it reproducible and comparable to other studies.

- Both methods are used in a single study, as the qualitative data helps to explain the quantitative data. The two sets of data are merged in the interpretation phase in order to form a comprehensive overview of post-edited quality.

- It adopts pragmatism as a worldview. It is problem-centred and real-world oriented, collecting data by using an approach that is informed by its practicality in a real-life application, at the example of the Norton Communities, with a view to facilitating an enhanced user experience and communication in the same.

4.3 Pilot Study

Based on what has been discussed in Part I of the Literature Review in Chapter 2, an initial methodology was developed to investigate the concept of community post-editing quality. In order to test this methodology for rigour, a pilot study was conducted in October 2012 with the purpose of answering the following research question:

Does monolingual/bilingual post-editing improve the quality of machine translated user-generated content in an online community environment?

The aim was to establish what kind of output community post-editors can produce in a bilingual and a monolingual set-up considering their knowledge of English and the Norton products, as explained in detail by Mitchell, Roturier and O'Brien (2013).

4.3.1 Set-Up

Ethical approval[2] was granted by the Dublin City University Ethics Committee prior to conducting the pilot study for both the pilot and the full-scale study. Due to limited resources, the participants for this study completed both monolingual and bilingual post-editing tasks.[3] Each task contained a subject line, the initial question that was posted in the forum and the post marked as the solution to the question. Four users were recruited for each language pair (EN-FR, EN-DE), with one participant (for EN-DE) completing monolingual tasks only.[4]

---

[2]see Appendix A

[3]Bilingual here means that post-editing was performed with access to the source text. The post-editors had some knowledge of the source language. However, they were predominantly not assumed to be bilinguals of German and English.

[4]This participant dropped out of the study after completing the monolingual tasks. This was beyond our control, as the participants were volunteers.

Both the pilot study and the full-scale study used the baseline ACCEPT MT system. It was trained on bilingual data both from in-domain data, e.g. product manuals of Symantec, and out-of-domain data, i.e. WMT12 releases of EUROPARL and news commentary (English-German, English-French). When training an SMT system, it is preferable to use a corpus that is close to the texts that will be translated with it (in-domain), i.e. in this context, domain specific texts. Out-of-domain data was used as supplementary data, to increase the coverage of out-of-domain terms and phrases, as it is not a purely technical domain. The ACCEPT SMT system was chosen for this research, as it was trained on data relevant to the domain and content to be post-edited (user generated content with a technological topic). More information on the system can be found in the ACCEPT deliverable D 4.1 Baseline MT systems.[5]

We avoided presenting participants with the same task more than once to avoid familiarisation and bias. A method of clustering similar tasks together was deployed (explained in section 4.4.5) to address this problem. Two sets of reference translations were created. One used a more formal register, while the other used more informal language. With these, we aimed to test whether the post-edited segments are closer to formal or informal language on the assumption that the language used in user-generated content would more closely approximate the informal reference language.

4.3.2 Procedure

The post-editors completed a short pre-task survey on demographics, their activity in the Norton forum and language skills. Subsequently, they completed the post-editing tasks in the ACCEPT portal (see Roturier, Mitchell and Silva 2013), with four

---

[5]http://www.accept.unige.ch/Products/D_4_1_Baseline_MT_systems.pdf

monolingual tasks, four monolingual tasks with translation options[6] and four bilingual tasks. It was evident from the data collected that the translation options were sparsely used, which led to the merging of the two monolingual designs. After the post-editing tasks, the content was evaluated manually (cf. section 4.3.3).

4.3.3   Evaluation

Both human evaluation and automatic metrics were used in the pilot study as quality measures. For human evaluation, adequacy and fluency (LDC 2005) and comprehensibility were rated on a 5-point Likert scale. Comprehensibility as measured in Roturier and Bensadoun (2011) was defined as "the extent to which the text as a whole is easy to understand [...] and the extent to which valid inferences can be drawn by combining information from different parts of the document" (Hovy, King and Popescu-Belis 2002). There was one evaluator per language-pair, a native speaker of the target language. As mentioned in Chapter 3 section 3.2.1, two automatic evaluation metrics were considered in this pilot study to identify whether they can give some indication on either the quality of the post-edited content or how quality and the number of changes made impact the quality of the post-edited content. Meteor 1.4 (Denkowski and Lavie 2011) and TER (Snover et al. 2006) were used to evaluate the post-edited content. While Meteor addresses weaknesses of IBM's BLEU metric by matching words based on the same stem and between synonyms (Banerjee and Lavie 2005) and was used with both sets of reference translations, TER calculates the number of edits that would need to be performed in order to change a machine translated output to match a reference

---

[6]Translation options are alternative translations (word/phrase-based) offered when clicking on high-lighted words.

translation semantically (Snover et al. 2006) and was used twice, comparing it against both sets of reference translations.

4.3.4 Results

For German, the results for all three criteria, fluency, comprehensibility and adequacy were better for bilingual post-editing than for monolingual post-editing. Nevertheless, a considerable number of segments were improved for monolingual post-editing, 43% (32 out of 75 segments) of segments, compared to 56% (16 out of 28 segments) of segments for bilingual post-editing. Furthermore, a larger percentage of segments improved through monolingual post-editing in terms of adequacy than through bilingual post-editing. The scores for adequacy did not approach 100%, which can be attributed to the fact that the post-editors may have had a limited knowledge of English. The average maximum scores have been found to be less than 100% in similar studies (e.g. de Almeida 2013, Koehn 2010).

In order to identify whether AEMs are a useful method for measuring community post-editing quality, the following needs to be considered. The results produced by the AEMs did not correlate with the human evaluations. In general, AEMs were not considered a reliable method for measuring quality, due to the small data sets available in the pilot and the full-scale study. AEMs revealed in this pilot study that post-editing solutions by participants who changed large chunks of the segments were penalised, while, on closer investigation, solutions that had been changed more than others performed better in the human evaluation. A more extensive discussion of the automatic metrics for the pilot study can be found in Mitchell, Roturier and O'Brien (2013).

4.3.5 Implications

Conclusions drawn from the experience gathered in the pilot study are the following. Firstly, recruitment was an unexpected challenge, as members of the German Norton Community were not responsive in large part and were hard to enthuse for the study. A more extensive recruitment strategy was devised for the full-scale study (cf. section 4.4.4).

Secondly, participants were frustrated by the quality of the machine translated output, which they expressed in requests to see the source text, for example. Some segments, at times a whole task, were left unedited; one participant dropped out of the study after the first eight (monolingual) tasks. For the full-scale study, pre-editing (see section 4.4.6) and the choice of post-editing bilingually was introduced (see section 4.4.10). Thirdly, there were some technical issues encountered, consisting of login problems, an unfamiliar and unintuitive user-interface and a help button that was not prominent enough. For the full-scale study, the user interface was simplified and a training video was created.

Lastly, AEMs were deemed to be inappropriate for measuring post-editing quality in a meaningful way, as there was a wide range of quality and styles observed among the post-editors, which led to low AEM scores.

The human evaluation for all three criteria seemed to be ambiguous especially for fluency and comprehensibility. It was not deemed necessary to include all three criteria. Since adequacy and fluency are the criteria that are most often used in previous studies and as comprehensibility is hard to distinguish from fluency, adequacy and fluency were kept in an attempt to reduce ambiguity. Adequacy was a crucial aspect because

the information is required to be transferred accurately in order to solve problems, while fluency is crucial in understanding that information quickly and to keep user frustration to a minimum. The data gathered prompted changes in the experimental design, presented below.

4.4 Full-Scale Study

This section outlines the final methodology in detail and how it was shaped by the knowledge gained from the pilot study. This involves presenting the research questions and variables involved, the experimental design and the methods employed to collect and evaluate the relevant data.

4.4.1 Research Questions

The motivation behind the full-scale experiment was to identify whether lay post-editors are able to post-edit MT content successfully and to identify any connection between the quality of the content and the post-editor profile. To investigate this in detail, three research questions were identified:

RQ1) *How does post-edited output produced by community post-editors compare to raw machine translated output?*

RQ2) *What are segment characteristics and post-editing patterns that render post-editing successful/unsuccessful?*

RQ3) *Do elements of the post-editor profile[7] serve as predictors for the quality of the post-edited output?*

---

[7]as described in section 4.4.9

In order to test the applicability of the TC model presented in Chapter 2, section 2.2.1.3 for community post-editing, the following research question was devised:

RQ4) *Is Göpferich's Translation Competence model applicable to community post-editing?*

RQ1 and RQ2 are of importance for identifying whether community post-editing is a feasible concept and an alternative to raw machine translated output, the results of which are presented in Chapter 5 and Chapter 7.

RQ1 implicitly deals with the concept of quality, which, as demonstrated in Chapter 3 section 3.2.2.5, is a concept that greatly depends on the translation/post-editing context. In an attempt to obtain a comprehensive picture of quality, it is measured both within and outside the community, employing an uncontrolled evaluation of fluency by community members, a controlled evaluation of fluency and adequacy by domain experts and an error categorisation performed by the present researcher (see section 4.4.12).

All hypotheses presented in this chapter are null hypotheses, which we are seeking to reject based on data gathered in the full-scale experiment:

H1.1$_0$ *Community post-editing produces content that is not of better quality than raw MT output as judged by community members, domain experts and/or the errors it contains.*

H1.2$_0$ *There are no correlations between the results of the community evaluation, the domain expert evaluation and the error annotation.*

H1.1 focusses on the quality produced by community post-editors compared to raw MT output. H1.2 investigates the relation between the three measures of quality and their appropriateness.

RQ1 tackles the concept of quality from a quantitative perspective evaluating the data gathered empirically. RQ2 then investigates post-editing quality on a more granular level from a qualitative perspective by investigating post-editing behaviour that produces high quality/ low quality post-edited segments (cf. section 4.2).

Chapter 7 initially focusses on RQ3. Despite the fact that previous studies have shown that the translator or post-editor profile, e.g. training or experience, are not correlated with the post-editing quality (de Almeida 2013), it is possible that characteristics of the lay post-editor profile, such as SL skills and domain knowledge, can serve as predictors for the post-editing quality, as it is outside a professional framework. It thus investigates the relation between the post-editor profile and the post-edited output, aiming at establishing which - if any - indicators, such as language skills, may be predictors for successful community post-editing. Some aspects of the post-editor profile are simultaneously sub-competences as described by Göpferich.[8]

H3.1$_0$ *There is no correlation between the reported attitudes towards machine translation and post-editing and the quality of the post-edited content.*

H3.2$_0$ *There is no correlation between the reported importance of grammar, content, spelling and punctuation and the quality of the post-edited content.*

---

[8]1. communicative competence in at least two languages, 2. domain competence, 3. tools and research competence, 4. translation routine activation competence, 5. psychomotor competence, 6. strategic competence

For H3.1, we hypothesise that the more positive the attitude towards machine translation and post-editing is, the better the post-edited content performs in the evaluation. Dealing with H3.2, we hypothesise that the more important these aspects are to the post-editors, the better the quality of the output, as a higher importance may indicate higher sensibility towards these types of errors and a higher chance that these are corrected.

The remaining hypotheses involve post-editor profile characteristics that represent the sub-competences as devised by Göpferich. As tools and research competence and translation routine activation are very specific to translation competence and training, they could not be measured directly here with the resources available and the time restrictions. It is believed that in order to gain understanding of a post-editors' strategic competence, an investigation on a more cognitive level, e.g. interviews or think-aloud protocols would need to be conducted, which would have been beyond the scope of this thesis. It is an aspect that could benefit from being investigated in the future, as it is thought to be the controlling and most important sub-competence in determining translation (or post-editing) quality (see Göpferich 2013, PACTE 2005).

$H3.3_0$ *There is no correlation between communicative competence in English and German[9] and the quality of the post-edited content.*

$H3.4_0$ *There is no correlation between the domain competence and the quality of the post-edited content.*

$H3.5_0$ *There is no correlation between the psychomotor competence and the quality of the post-edited content.*

---

[9]This is measured here by reported language competence (English and German) and the quality of the post-edited content.

In regards to H3.3 we hypothesise that the higher the reported language competence (English and German), the better the post-edited content performs in the evaluation. This is based on the assumption, also represented in Göpferich's model, that a better communicative competence in these two languages leads to better post-editing competence. For H3.4, the more domain knowledge the post-editors have, as measured by user statistics in the communities,[10] the better the post-edited content performs in the evaluation. It is assumed that domain competence leads to better post-editing competence. Psychomotor competence is measured in the number of keystrokes and post-editing (typing) time spent per task. In reference to H3.5, we hypothesise that a higher number of keystrokes and time spent post-editing, the better the resulting PE quality. This is based on observations made in the pilot study were high-scoring segments had often been changed greatly.

To operationalise RQ4, then, the insight gained from RQ3 will be used to critically discuss Göpferich's model of Translation Competence, the methodology used in this study to measure the competence and the transferability of this model to post-editing by lay translators.

In summary, Chapter 5 aims at establishing the quality of the community post-edited content on a quantitative level, Chapter 6 on a qualitative level, while Chapter 7 investigates the relationships between quality and the post-editor profile and attitudes and takes a first step towards testing the applicability of Göpferich's translation competence model to post-editing by community post-editors.

---

[10]With these user statistics, we hoped to reduce subjectivity as introduced by self-reporting.

4.4.2 Variables

After having presented the research approach taken, the research questions that this thesis seeks to answer and the variables involved, Table 4.1 gives an overview of how the methods are mapped to the research questions and what research instruments are used to collect what type of data.

Research Question	Research Instrument	Data Type	Method
RQ1	•error annotation •domain specialist evaluation (fluency, adequacy) community evaluation (fluency)	numerical data	QUAN
RQ2	•ST segment characteristics (e.g. segment length) •TT segment characteristics (e.g. number of errors)	numerical data	QUAN
	•TT segment characteristics	post-edited examples	QUAL
RQ3	•post-editors' profile •closed-ended survey questions	numerical data	QUAN
RQ4	•linking sub-competences to theoretical data	numerical data	QUAL

TABLE 4.1: Overview of methods to answer RQs

In order to test the hypotheses for RQ3 (and RQ4), the following variables were identified, as summarised in Table 4.2:

Dependent	Independent	Hypothesis
quality	•reported attitude towards machine translation/ post-editing	H3.2
	•reported attitude towards content, grammar, spelling and punctuation	H3.1
	•reported language competence	H3.3
	•reported domain knowledge	H3.4
	•number of keystrokes & post-editing time	H3.5

TABLE 4.2: Independent and dependent variables

The dependent variable to be investigated is the quality of the post-edited output, which is potentially influenced by the independent variables, e.g. self-reported language competence.

As explained in Section 4.2 and as can be seen from Table 4.1, the emphasis in this experiment was on the collection of quantitative data. This requires a statistically rigorous approach to analyse and interpret the results. The data collected is aggregated statistically (e.g. presenting means, standard deviations). Relations between pairs of variables, such as language competence (independent) and post-edited quality (dependent) variables are investigated by calculating Spearman's correlation.[11] In addition, the proportion of difference between two sets of data, e.g. fluency scores by domain specialists and fluency scores by community-members are calculated in addition to correlations, as results can be correlated in that they move in the same direction given specific circumstances but one can be significantly lower than the other. Both proportion of difference and correlations are tested for statistical significance, dismissing the possibility of the results obtained having occurred at random. This approach is taken for RQ1, RQ3 and RQ4. The statistical results are calculated using Stata and SPSS.

4.4.3 Experiment Procedure

After having established the theoretical basis for the full-scale experiment, this section presents the experimental design and procedure, followed by the procedures for data analysis. The experiment involved eighteen participants (cf. recruitment in section 4.4.4) editing 12 tasks.[12] For an explanation on the selection and pre-editing of material see sections 4.4.5 and 4.4.6 in the ACCEPT portal, as introduced in Chapter 1 section 1.1. They first completed a pre-task survey (section 4.4.8, Appendix B), which included a training video as developed by the present researcher[13] to circumvent

---

[11]Spearman's correlation coefficient is used due to a small sample size.

[12]One task consists of a subject line, a question asked in the community and the answer marked as a solution.

[13]http://www.youtube.com/watch?v=7aQLBL-gSao

technical difficulties.[14] Subsequently, the participants were able to access all 12 tasks on their home page of the ACCEPT portal. The participants edited the tasks in the ACCEPT portal using the editor and the guidelines described in section 4.4.10. Post-editing actions were recorded using an XLIFF format (see section 4.4.11) devised for the ACCEPT portal. After finishing all tasks, the community post-editors completed a post-task survey (cf. Appendix C).

4.4.4 Sampling

The population considered for the experiment comprised of visitors to the German Norton forum who are German native speakers. For the recruitment, users of the German Norton Community were identified with any kind of rank within the online community and any number of posts authored between 10 and 20,000, who were at least last active in December 2012. Thirty-eight private messages were sent out in a first round of recruitment and 53 in a second round, accompanied by a public post for recruitment and a post in the super-user board. This sampling strategy is non-probabilistic, as it focussed on users with a certain activity in the forum and profile, as described above. This more targeted approach to sampling was required, as a broad recruitment strategy had failed, as demonstrated in the pilot study (cf. section 4.3). Eighteen participants were ultimately recruited, including two non-native speakers of German who had expressed great interest in taking part in the study.[15]

---

[14]The researcher had no control over whether the participants watched the video or not.

[15]While the experiment had initially been restricted to native speakers of the TL, two non-native speakers of German were included, as they were highly motivated and they facilitated insight into post-editing in a non-native language.

4.4.5 Post Selection

The results of the post-editing experiments can only be relevant if the tasks used are representative of the body of texts it has been extracted from. The first decision to be taken was whether to use a test-suite or a corpus. While test suites are widely used to isolate individual linguistic features and investigate them (e.g. Balkan et al. 1994), a corpus-based approach allows for a more natural and data-driven - a pragmatic - approach, as occurrences that have been taken unaltered from real-life data can be examined (e.g. McEnery, Xiao and Tono 2006). Based on the fact that the focus is on translating user-generated content and on ensuring generalisability (across the Norton Communities for the language pair English-German), a corpus-based approach was taken. This decision also conforms with data selection strategies for test suites that are used in the development of machine translation engines (cf. Callison-Burch et al. 2012).

Three aspects that should be considered during corpus development are: stasis/dynamism, representativeness/balance and size (Kennedy 1998, pp. 60-70). For this experiment, a static rather than a dynamic corpus was accumulated, as the focus is on language rather than mass. It is believed to be difficult to ensure that conclusions drawn from a corpus-based approach can be extrapolated to the genre studied (e.g. Kennedy 1998). Bowker and Pearson, try to pre-empt non-validity by, for example, including control aspects, such as text length, subject and publication date (Bowker and Pearson 2002). As those aspects, especially text length and style, depend on the author, which is an unknown variable in an online support forum, this approach was discarded. For the full-scale experiment, all content (96,229 threads) from the English-speaking Norton Community from boards which discuss Norton products and contain

more than 500 posts was extracted between the time span of 1 January 2012 to 20 March 2013.

In order to increase the content covered in the post-editing tasks, rather than 12 tasks edited by all post-editors, two similar groups of 12 tasks were selected, thereby doubling the number of tasks. The post-editors were randomly split into Group A and Group B. To facilitate the formation of two groups of tasks that are quite similar in complexity, rather than selecting tasks randomly, a method of clustering was employed. Characteristics considered in this clustering technique were the following: number of words per post, number of sentences per post, Type-Token-Ratio of the post, number of normalisation tags (maskable tokens) ($< url_ph > | < path_ph >$ etc.),[16] and perplexity with respect to the full domain forum-based language model (LM). The forum-based language model is a 5-gram LM with modified Kneser-Ney (Kneser and Ney 1995) smoothing trained on the available monolingual English forum data (approximately one million sentences). It was trained using the IRSTLM language modelling toolkit (Federico, Bertoldi and Cettolo 2008). An unsupervised clustering approach based on the K-mean clustering approach (MacQueen 1967) and more specifically the open source K-Means algorithm in the Weka Toolkit were used for automatic clustering. The K-means clustering approach aims to group n observations into k groups to assign each observation to a group with the nearest mean.

The 96,229 threads collected contained 2,924 question and accepted solution pairs (= a task). 16 clusters were built using the same technique as in the pilot study, i.e. with the idea of selecting two tasks from each cluster (one for group A and one for group B) and having 4 spare clusters in case a cluster was not deemed suitable. Eight posts

---

[16]These tags were inserted in a cleaning step, in which structures, such as URLs were replaced by tags.

were randomly selected from each cluster (= 128 tasks). This number was reduced by inspection of these tasks and discarding unsuitable clusters and tasks. Tasks were inspected manually and deemed unsuitable if the accepted solution did not actually contain the solution to the problem, or if the length of a task differed greatly from the length of the other tasks in the cluster etc. As only a small percentage of the forum content (0.17%) was represented in the post-editing task, it was of importance to select content that would be useful when post-edited. Two tasks were selected from each cluster that appeared to be the closest to each other.

With this approach, a total of 4,088 words (machine translated) were presented to the post-editors split into two groups, i.e. each post-editor edited 2,042 (Group A) or 2,046 (Group B) words.[17] While the number of words comprises a small portion of the forum content available, it is greater than samples usually used in post-editing studies thus far (e.g. 1,008 words in de Almeida 2013; 1,025 and 1,101 words in Koehn 2010).

4.4.6 Pre-Editing

Basic pre-editing was deemed an essential step in the experiment procedure, based on participant feedback from the pilot study, which indicated that the raw MT output was of too low a quality. The results for the language pair English-German continue to be one of the lowest in the development of statistical machine translation (e.g. Bojar et al. 2014). While English and German are closely related languages, they differ considerably in their syntax. Especially long-range reordering of verbs that is required to produce fluent German output continues to pose a challenge for machine translation in general and statistical machine translation systems in particular. In

---

[17]The source texts can be found in Appendix D and the raw MT output can be found in Appendix E.

addition, problems that naturally arise in user generated content, such as posts in the

Norton Community, supported the inclusion of pre-editing. An excerpt of a thread in

the Norton Community is presented below:

> 360 problem's
> hi everyone first time poster hopefully last time aswell i woke up this morn-
> ing started up my desktop and went to run norton 360 it wouldnt open so
> i clicked it again still wouldnt open so i did a fresh reinstall chose option 1
> restarted pc then got it running did a full system scan updates made me
> restart my pc so i did came back into windows hit the icon wouldnt start
> up then went to uninstall then froze had to end process through windows
> task manager been at it for 4hrs now it was working fine yesterday or the
> day before i am running win7 ultimate 64bit
>
> ps. uninstalled again with option2 full uninstall its just not responding
> at all just hangs when i hit the norton icon
>
> ——— Hi *username*,
> Sorry for the delay in replying - I am located in the colonies and at the
> time you posted (all three of your posts) I was in bed - even volunteers here
> at Norton are permitted to sleep! ;-) [...][18]

This excerpt demonstrates some of the problems found in UGC, ranging from grammar

issues, e.g. "360 problem's", to emoticons, to missing punctuation and the unrelenting

lower casing of all words. These render the text hard to read in general and extremely

difficult to translate correctly for an MT system in particular. The mix of technical

terms and colloquialisms pose additional problems to any machine translation system.

While not all of these issues can and should be resolved by pre-editing - as it may alter

the style and sentiment of the source text - a subset of the ACCEPT pre-editing rule

set was implemented manually. The selection of rules was informed by an analysis of

the lowest scoring segments from the pilot study. It emerged that the most prevalent

problems were: wrong verb/verb part missing (present in the German MT output),

wrong terminology (MT), colloquialism (present in the English ST), sentence longer

---

[18]https://community.norton.com/forums/360-problems

than 30 words (ST), missing (end) punctuation (ST) and spelling (ST). The first three issues were disregarded, as they focus on problems in the MT engine, would alter the style greatly or would not be reproducible due to multiple alternative translation options. Sentence length and punctuation, as well as spelling rules were employed:[19]

1. Sentence too long (longer than 30 words) – shorten sentences in as few edits as possible
2. Sentence ending punctuation – insert missing punctuation marks at the end of sentences
3. Spelling error (as indicated by ACCEPT pre-editing tool)

This decision is in line with similar research in the field. For example, O'Brien and Roturier (2007) find that rules dealing with spelling errors, incorrect punctuation and long sentences (more than 25 words) had a significant impact on post-editing effort and comprehensibility. Tatsumi et al. suggest that "longer source sentences may cause difficulty for CPE contributors to produce acceptable quality" (2012, p. 8).

The manipulations to the ST employed in this experiment, which may affect the post-editing results, are easily reproducible. In the future, it would be hoped that these rules would be either implemented automatically or that a pre-editing facility would be available in a non-intrusive way to the users of the Norton forums.

4.4.7   Machine Translation

Following the pre-editing step, the content was machine translated using the statistical MT system based on Moses (Koehn et al. 2007) produced within the framework of the ACCEPT project. It was the same system as employed in the pilot study and described in section 4.3.1. The machine translated texts used as tasks in this experiment can be found in Appendix E.

---

[19]For detailed figures on the impact of these rules see ACCEPT (2013b, p. 7)

4.4.8 Pre-Task Survey

In order to gain informed consent and information on their profile from the participants, they completed a pre-task survey on the online survey platform Surveymonkey[20] prior to the post-editing tasks.

The survey was designed by grouping together questions on the same topic and breaking them into pages, starting with general questions and moving to more project specific questions as the survey progressed. The first page focussed on general information about the user. In order to get a brief demographic overview of the sample that was recruited from the Norton Community, the participants were asked to indicate their gender and their age category. For both questions, the option of "I do not wish to specify" was available to avoid participants skipping questions or dropping out of the study. The survey then focussed on the participants' attitudes towards different linguistic aspects in the forum and attitudes towards MT and PE. The questions are described in more detail in the following section, with a full list of questions presented in Appendix B.

4.4.9 Post-Editor Profiling

In order to answer RQ3 and RQ4[21] the post-editor profile needs to be defined. It was created using a pre-task survey (cf. Appendix B), a post-task survey (cf. Appendix C), user statistics gathered from the Norton Community and statistics gathered from the post-editing process. Aspects of the post-editor profile are independent variables that were hypothesised to have an effect on the post-editing quality. Here, the post-editor

---

[20]http://www.surveymonkey.com
[21]Does the post-editor profile influence the quality of the post-edited output? Is Göpferich's Translation Competence model applicable to community post-editing?

84

profile consists of the following aspects: attitudes towards types of errors, and attitude to MT/PE; and according to Göpferich's model: communicative competence (German and English), domain competence and psychomotor competence.

The first component of the post-editor profile is their attitude to machine translation and their self-reported motivation to post-edit. A 5-point Likert scale is used here, as it is a common scale in comparable recent research to measure attitudes (de Almeida 2013, Doherty and O'Brien 2012). This data was captured prompting post-editors to indicate their agreement with the following two statements on a 5-point Likert scale, where 1 indicates "strongly disagree" and 5 indicates "strongly agree".

The second component of the post-editor profile is the attitude towards the following aspects in the Norton community: grammar, content, spelling and punctuation.[22] The participants were asked to rate the importance of these in the Norton forum on a 5-point Likert scale, with 1 indicating "not important" and 5 indicating "important" (cf. de Almeida 2013). The order of these items (grammar, content etc.) was randomised for each user to avoid bias of ranking.

1) Machine Translation may be useful in the Norton Community.
2) I am motivated to edit machine translated posts.

The first aspect of Göpferich's TC model is language competence, here split into English language competence and German language competence and forms the third component of the post-editor profile. Language skills could not be tested directly, e.g. with a language test (reading/listening/speaking), due to the participants being remote and having joined the experiment on a voluntary basis and the desire to keep drop-out rates to a minimum. Thus, language competence for both German and English was

---

[22]It was initially considered to gain information on errors on a more granular level. This was dropped due to low numbers in the individual error categories of the machine translated output, which would have prevented a sound empirical investigation.

measured based on the self-assessment grid for European languages developed by the European Centre for the Development of Vocational Training (European Centre for the Development of Vocational Training CEDEFOP 2011). For German, the descriptions from the "Writing" category were employed, as post-editing is about correcting texts in writing. It was hoped that the reported German language competence would give some indication of the linguistic preferences described below:[23]

> Please rate your German skills (WRITING). Indicate which category you feel you belong to the most.
>
> - A1 I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.
> - A2 I can write short, simple notes and messages. I can write a very simple personal letter, for example thanking someone for something.
> - B1 I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.
> - B2 I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.
> - C1 I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select a style appropriate to the reader in mind.
> - C2 I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

For English, the descriptions of categories were selected from the "Reading" section, as reading and understanding the original English text is a standard requirement for traditional bilingual post-editing.

---

[23]The original German version of the following questions, which was presented to the post-editors, can be found in Appendix B.

Please rate your English skills (READING). Indicate which category you feel you belong to the most.

- A1 I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.
- A2 I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.
- B1 I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
- B2 I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.
- C1 I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.
- C2 I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

The fourth component of the post-editor profile, the second sub-competence in Göpferich's model, is domain competence or domain knowledge, which is based on user statistics from the Norton forum, i.e. membership in the German and English forum, their assigned rank and the number of posts read and posted. These were believed to be indicators of domain competence, as a high number of posts written indicates that the user answers posts, rather than just posting questions. Typically, a once-off user asks one question and returns to answer within that same thread a few times (≤ 10 times) and never returns after that. Furthermore, we assume that the more posts the user has read, the more knowledge they gain on the Norton products and solving problems related to them. In addition, ranking information was used cautiously, as the ranking pattern is not always consistent. Due to this, participants were split into three rank groups (guru, Symantec employees and other), as presented in Chapter 7, section

7.5. This information was retrieved from the section of the user profiles in the Norton Community that is only visible to community administrators. Kudos was not used as a measure, as it appeared to coincide with guru status predominantly.

Psychomotor competence is the last component of the post-editor profile and the fifth sub-competence in Göpferich's model - the final one to be considered for the purpose of this thesis. Data for this sub-competence was collected during the post-editing process, in the form of number of keystrokes and the time spent post-editing.

4.4.10 Interface

After having outlined the preparatory steps prior to the post-editing tasks, this chapter now introduces the post-editing platform and the post-edited data collection process.

The participants completed the post-editing tasks in the ACCEPT portal. The post-editing interface is displayed in Figure 4.1. The left half of the window shows the full text to be edited for that particular task. In the top right box, the original segment is displayed. This can be changed from shown to hidden using the switch displayed above it, a feature developed based on the findings from the pilot study. In the edit box, the user can edit the current segment. Comments are saved in the edit box at the bottom right. All edits are saved automatically. Further functionalities are the undo/redo buttons and the ACCEPT spelling and grammar checker.

Post-editing guidelines were available to the participants through the ACCEPT post-editing interface. They were developed based on the following assumptions. Firstly, post-editors in the Norton Community are not expected to be professional translators/linguists. Secondly, the participants are lay post-editors, which means that providing
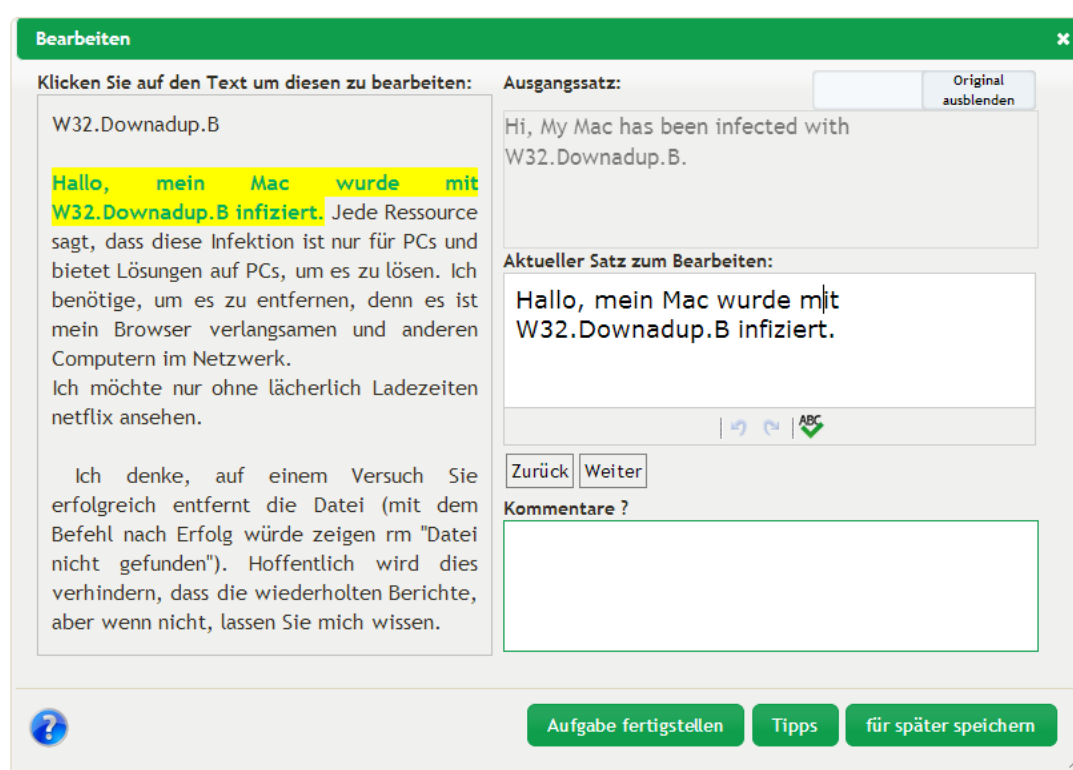
FIGURE 4.1: Post-editing interface

them with the same comprehensive guidelines that translators receive may be inappropriate. Usually, corporate post-editing guidelines are developed and used in-house and not available to the public (cf. de Almeida 2013, p. 38). There are different sets of guidelines for bilingual post-editing available for industrial (e.g. GALE Post-editing Guidelines 2007) and academic use (e.g. de Almeida 2013, Tatsumi et al. 2012). Furthermore, post-editing guidelines have been developed seeking to establish a standard (TAUS 2010). However, these guidelines have been developed for translators or at least people with advanced linguistic skills. The guidelines developed for the ACCEPT portal were kept as concise and comprehensible as possible. They were largely based on the TAUS Post-Editing Guidelines for "good enough" quality (TAUS 2010) and combined with the monolingual post-editing instructions as presented in Koponen (2011):

- Aim for semantically correct translation.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling apply.
- No need to implement corrections that are of a stylistic nature only.
- No need to restructure sentences solely to improve the natural flow of the text. (TAUS 2010 [Online])

"Test subjects were instructed to edit the text (based on raw MT only) into fluent and clear Finnish according to how they interpret the meaning" (Koponen 2011 [Online]).

This approach was taken, as TAUS intended their guidelines to be used as a starting point for developing PE guidelines relevant and customised to one's own settings (TAUS 2010). For this study, the goal was to make them appropriate for volunteer post-editors with potentially no experience in translation/linguistics that are suitable also for monolingual post-editors. The guidelines were reduced to four items. The heading was changed to "tips" in German rather than guidelines, to make them more appropriate for the volunteer context. The guidelines were displayed by clicking the "Tipps" button (cf. Figure 4.1) and were as follows:[24]

Tips for post-editing:

- Edit the text to make it more fluent and clearer based on your interpretation.
- Try to correct phrasal ordering and spelling, for example, if they make the text hard or impossible to understand.
- Use words, phrases or punctuation as they are, if they are acceptable.
- If you are working with reference to the original text, make sure that no information has been added or deleted.

---

[24]The German guidelines are available in Appendix F.

4.4.11   Data Collection

To make the tasks available to the post-editors, two projects (Group A and Group B) were created in the ACCEPT portal, containing 12 tasks each. Each task had been formatted in a JSON format,[25] as required by the portal. Besides an array of source segments and matching target segments, this format allowed a manipulation of how the raw machine translated output is presented to the post-editor, i.e. creating paragraphs. A comprehensive overview of the project creation options and JSON file format is presented in Roturier, Mitchell and Silva (2013). Most actions performed by the post-editor in the post-editing environment were recorded using the ACCEPT API in ACCEPT reports. These reports can be retrieved and displayed by project administrators in the form of an XLIFF 1.2[26] compliant file in real time. The structure of the XLIFF document is based on the reports developed by Specia et al. (Aziz, de Sousa and Specia 2012). One task is mapped to one element in the XLIFF file, which contains a *header* and a *body* element. The *header* element contains meta data on the text that has been produced, such as keystrokes, the switches to show or hide the source text[27] and the post-editing time. The *body* element contains the actual text of a segment. For each segment, the text of the source sentence is recorded, the target sentence (the final version of the sentence), and any alternative translations. Alternative translations can be previous revisions of the current segment or the machine translated segment, if the user changes the segment once and does not revise it again. The target element contains the last revision of the segment, below that, the original

---

[25] http://json.org/

[26] http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html

[27] Whenever a participant opened a task, the switch was set to the project default, here hiding the source. We were hoping to record when the participants choose to see the source, i.e. when do they feel that the monolingual raw MT output is not adequate enough to post-edit. Apart from one participant, all participants switched to see the source text for all tasks, regardless of their English competence.

machine translated segment is presented as the first alternative translation, sequentially followed by any other revisions. Metadata, such as a comment on a segment, are mapped to the respective revision of a segment by using unique phase names. An example of an XLIFF file is presented below in Figure 4.2.

```xml
'<xliff version="1.2">
 ▼<file original="69fb55d63b6636a66579b864e03d63f9c6fc0c1351624e8db98553dd884eeccb912b02a550e31
   language="en" target-language="de" datatype="" category="" product-name="">
   ▼<header>
     ▶<phase-group>...</phase-group>
     ▼<count-group name="1">
        <count phase-name="r1.1" count-type="x-editing-time" unit="x-seconds">10.328</count>
        <count phase-name="r1.1" count-type="x-start_source_switch" unit="instance">show</count>
        <count phase-name="r1.1" count-type="x-source_switch" unit="instance">2</count>
        <count phase-name="r1.1" count-type="x-source_switch" unit="x-seconds">17.808</count>
        <count phase-name="r1.1" count-type="x-source_switch" unit="x-seconds">8.682</count>
      </count-group>
     ▶<count-group name="2">...</count-group>
     ▶<count-group name="3">...</count-group>
     ▶<count-group name="4">...</count-group>
     ▶<count-group name="5">...</count-group>
     ▶<count-group name="6">...</count-group>
     ▶<count-group name="7">...</count-group>
     ▶<count-group name="total">...</count-group>
   </header>
   ▼<body>
     ▼<trans-unit id="1">
        <source>IS this a virus</source>
        <target phase-name="r1.1">Ist dies ein Virus</target>
       ▼<alt-trans phase-name="mt_baseline">
          <target>Ist dies ein Virus</target>
        </alt-trans>
      </trans-unit>
     ▼<trans-unit id="2">
       ▼<source>
          An Icon shows on my computer and it is rminstall[1].exe.
        </source>
       ▼<target phase-name="r2.1">
          Zeigt ein Symbol auf meinem Computer, es ist die rminstall [1] .exe
        </target>
       ▼<alt-trans phase-name="mt_baseline">
         ▼<target>
            Zeigt ein Symbol auf meinem Computer, und es ist rminstall [1] .exe
          </target>
        </alt-trans>
      </trans-unit>
     ▼<trans-unit id="3">
        <source>I cannot find a file for this on drive c.</source>
        <target phase-name="r3.1">Ich kann dafür keine Datei auf Laufwerk c finden</target>
       ▼<alt-trans phase-name="mt_baseline">
         ▼<target>
            Ich eine Datei nicht finden kann für diese auf Laufwerk c.
          </target>
        </alt-trans>
     ▶<trans-unit id="3">...</trans-unit>
     ▶<trans-unit id="4">...</trans-unit>
     ▶<trans-unit id="5">...</trans-unit>
     ▶<trans-unit id="6">...</trans-unit>
     ▶<trans-unit id="7">...</trans-unit>
   </body>
 </file>
```

FIGURE 4.2: Sample XLIFF report

4.4.12   Evaluation

Subsequent to the completion of the post-editing tasks by the participants, the data was extracted from the XLIFF reports for further analysis. The files were processed automatically with Python scripts, extracting and aggregating the data required to test the hypotheses.

Quality of the post-edited content is the dependent variable and is the central concept for RQ1 and RQ3.[28]   As discussed in Chapter 2, there is no absolute quality for translated or post-edited texts, as the concept of quality is subjective and needs to be defined according to its context. Here, it was not only important to ensure that the translation process was successful both in terms of meaning and language from a domain specialist and a linguistic perspective but that the end-users were satisfied with the translations. End-users in this context were members of the community who may value a quick solution to their problem, i.e. content, over a slower but grammatically correct and otherwise immaculate solution. These requirements informed the design stage of the evaluation, as presented below.

A commonly used approach to evaluating (human) translation quality is that developed by House (House 1977, House 1997, House 2015). She approaches the underlying analysis of ST and TT on a macro level, e.g. genre and register, and a micro level, e.g. lexical means and textual means. She focusses especially on the language pair German-English, which is very relevant for this thesis. However, the text domain and author, and the 'commission' of the translation are not an ideal fit for her approach of

---

[28]How does post-edited output produced by community post-editors compare to raw machine translated output? Does the post-editor profile influence the quality of the post-edited output?

predominantly professionally authored high-quality texts, as this thesis deals with user-generated forum content. Hence, we selected a quality evaluation approach that seemed to be better suited to the texts (STs and TTs) at hand, here the error annotation developed by de Almeida (2013) to tag post-edited content, as described in section 4.4.12.1.

4.4.12.1 Error Annotation

The first type of evaluation was performed by the present researcher and consisted of an error annotation. This type of evaluation constitutes a top-down approach as presented by Drugan (2013), a quality evaluation approach widely employed especially in the translation industry. It facilitates an assessment of the quality produced in an online community environment compared to that of professionally translated content in the industry to a certain extent. The error annotation was based on the error categorisation developed by de Almeida 2013.

The reasoning behind selecting de Almeida's typology is the following. It is a typology that is useful for the present experiment which, firstly, allowed for the classification of both machine translated and post-edited text; secondly, for user-generated content; and thirdly, for content produced by post-editors untrained in linguistics. Previously, error typologies had been developed for classifying errors in either MT output (e.g. Vilar et al. 2006) or post-edited output (de Almeida 2013, Koponen 2010). While Vilar's typology is widely used, it is not suitable here, as it categorises errors in terms of state, i.e. missing or incorrect, rather than categorising by error type, e.g. grammar, sense, spelling etc. Koponen's typology focusses on categorising errors in post-edited content with categories of concept, participant (agent) and relation. Her sub-categories

are omission and mistranslated, which reflect the focus on post-edited content rather than machine translated content. However, it does not distinguish between grammar and spelling and can be difficult to apply to PE data due to ambiguity. De Almeida develops a typology specifically for post-edited content She bases it on the LISA QA model (LISA 2009), which is no longer maintained but has been widely used in the localisation industry. The LISA QA model's categories include accuracy, mistranslation, terminology, language, developed to identify errors in (human) translations as a quality assurance step. This human component is what makes it interesting to this research. De Almeida's typology combines aspects of the LISA QA Model, GALE post-editing guidelines and additional categories (cf. de Almeida 2013, pp. 86–96):

Master categories			
Essential changes	Preferential changes	Essential changes not implemented	Introduced errors

Subclassification for the four master categories	
Main Categories	Subcategories
Accuracy(completeness)	Extra information in MT output
	Information missing from MT output
	Untranslated text
Consistency	N/A
Country	Decimal points
	Quotation marks
	Currency symbol
	Date/time format
Format	N/A
Language	Adjectives
	Adverbs
	Capitalisation
	Conjunctions
	Determiners
	Gender
	Nouns
	Number
	Phrasal Ordering
	Prepositions
	Pronouns
	Punctuation
	Spelling
	Verb tense
Mistranslation	N/A
Style	N/A
Lexical Choice	N/A

TABLE 4.3: Error typology (de Almeida 2013, p. 95, reproduced from Table 3.6)

Colour coding:
Black – categories from the LISA QA Model
Blue – categories from the GALE PE Guidelines
Magenta – categories added by de Almeida 2013

The presented error categorisation was used for the purpose of identifying errors in the post-edited content.[29] As described in Mitchell, O'Brien and Roturier (2014), the categories Country, Consistency and Style were not considered in the experiment here. While there were no occurrences for the first category, the latter two were not applicable to forum content in general, where consistency in terminology and style are

---

[29]Throughout this thesis, the terms 'Accuracy', 'Language' and 'Format' are used capitalised when they refer to de Almeida's typology.

uncommon. Furthermore, the tasks for the domain specialist evaluation were composed of a set of four randomly picked post-edited versions of a pool of nine, thus rendering consistency impossible by default. The categories retained were accuracy errors (additional information, missing information, untranslated information, mistranslated information),[30] language errors consisting of any kind of grammar or spelling errors and format errors (here reduced to missing/additional spaces in order not to over-penalise "errors" that are due the nature of user-generated content).

A sample of 44% of all post-edited content collected rather than the whole pool of segments was annotated, as is common practice with error annotations. The content that was error annotated was the same that was evaluated in the domain specialist evaluation. Segments were extracted randomly with four post-edited versions for the source segments and one machine translated version. It was ensured that an equal number of segments per post-editor was chosen.

The tool used for the error annotation was BLAST (BiLingual Annotation/Annotator/Analysis Support Tool), a standalone tool for the error analysis of MT output developed by Stymne (2011). As BLAST may be used with any error typology, a typology corresponding to the de Almeida categorisation was encoded in a file that is readable by BLAST. The intuitiveness, simplicity and flexibility of the program informed the choice to use it here, rather than other tools that are equally effective, such as Appraise (Federmann 2010) but which would have involved more extensive effort in preparing and maintaining the error annotation.

Figure 4.3 displays the user interface of BLAST. The top half of the window displays three segments: a source segment, a corresponding reference segment (or machine

---

[30]De Almeida keeps this category separate. Here, it was integrated into the Accuracy category.

translated segment) and a machine translation system segment (or post-edited segment). The bottom half displays the available error categories, based on the category file used, which can be selected for the active word or phrase.
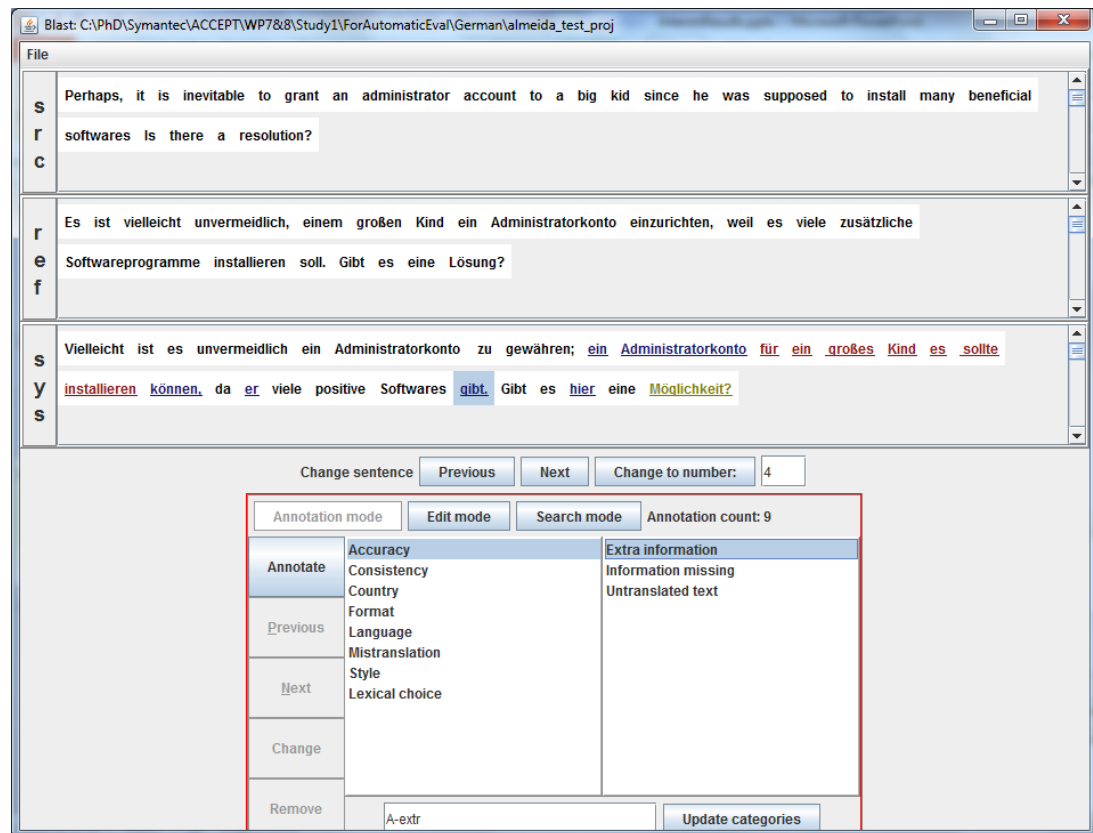
FIGURE 4.3: Screenshot of BLAST

Once the content is annotated for errors, BLAST automatically calculates the number of errors and the number of occurrences of each error type per file. All additional information was extracted from the respective BLAST project files. The raw machine translated output and 44% of the post-edited content (the same content that had been evaluated by domain specialists – see section 4.4.12.2) was annotated using this approach. The errors were not annotated in reference to a gold-standard reference. Rather, the smallest number of errors were identified that would have needed to be corrected to achieve a correct translation.

4.4.12.2   Domain Specialist Evaluation

Human evaluation was believed to be better suited for the evaluation of post-edited content produced by volunteers of the Norton Community than AEMs as confirmed in the pilot study (cf. section 4.3). Based on observations made during the pilot study, it was also anticipated that the post-editing steps and approaches would be unpredictable, due to varying language skills and profiles of the volunteer post-editors.

Drawbacks to this approach are low inter-annotator and intra-annotator agreement, subjectivity and the difficulty of judging both concepts consistently for long sentences (Denkowski and Lavie 2010). While long sentences were eliminated in the pre-editing step of this study, we attempted to pre-empt the agreement problems by informing the evaluators about the evaluation guidelines and expectations in face-to-face meetings and using a training video. Relative ranking, the solution suggested by Denkowski and Lavie to decrease subjectivity, does not seem adequate here. The post-editing strategies employed by community post-editors are unpredictable and most likely to differ greatly. While relative ranking facilitates the identification of the best/worst MT systems or post-editors, the rating of adequacy and fluency gives insight into the individual solutions by the post-editors independently and relative to each other. Thus, the approach chosen for the domain specialist evaluation is that of judging both adequacy and fluency on a 5-point Likert scale.

Adequacy was rated on a five-point Likert scale adopted from LDC (2005). It was rated for each segment in comparison to a gold-standard reference translation, in a monolingual manner. It has been discussed whether gold-standard translations produced by professional translators are always an adequate choice. We deemed it a reasonable

choice here, as knowledge of the community post-editors' profile was not available and the translation quality was unpredictable. Furthermore, the gold-standard reference translations were verified by the present researcher.

The concept of fluency was also adopted from LDC (2005) and rated on a five-point Likert scale. A minimum of two evaluations for adequacy and fluency (LDC 2005) were considered adequate, a minimum of three evaluators are considered essential by Dyson and Hannah (1987, p. 166), while a minimum of four evaluators to achieve reliable results is recommended by Arnold (1994, p. 162). The aim was to recruit eight evaluators. Due to a resource shortage, however, seven domain specialists,[31] three for Group A and four for Group B, were recruited internally. Regarding the evaluator skill set, domain knowledge was given priority over translation or post-editing experience, as the end-users of the translated content are users of the Norton Community and we assume their profile constitutes a closer match to that of the evaluators.

Prior to the evaluation, an introductory email describing the project and tasks was sent to the evaluators including the informed consent and plain language statement.[32] It contained a link to a training video developed by the present researcher demonstrating the evaluation for example segments and how to deal with ambiguous cases, as based on problems identified by Denkowski and Lavie (2010). It also contained a summary of the evaluation guidelines discussed in the video, which is presented below:[33]

- Adequacy and fluency should decrease with an increasing number of mistakes.
- BUT: if the meaning of just one word is wrong and this leads to a meaning opposite to the intended being attributed to the whole sentence, then adequacy should decrease dramatically

---

[31] A domain specialist is defined as a (former) Symantec employee working either in the localisation or the technological support department.

[32] They agreed to these in the Evaluator Survey, as can be found in Appendix G.

[33] The German translation can be found in Appendix H.

- Missing information should be penalised according to quantity.

- Additional information should be penalised according to how much the meaning of the reference translation has changed.

- Based on your domain knowledge, variation in the translations should be tolerated if the meaning of the sentence is still correct.

- Adopting an English term in German, as well as the translation of the same term should be equally tolerated.

- Do not spend more than a few minutes on the evaluation of a sentence and its corresponding translation candidates.

- Close a task if you are not currently working on it. (The time of the evaluation will be recorded.)

The evaluation platform used was the open-source tool Appraise, developed and described in Federmann (2010). This tool has been used in similar tasks, before, for example in the WMT shared tasks (cf. Callison-Burch et al. 2012a). Other tools, such as Amazon's Mechanical Turk were also considered but dismissed, due to lack of control over the evaluator profile. The code for Appraise, which is available on GitHub,[34] was modified for the purposes of the evaluation and an individual instance of the tool was set up.

The content evaluated in the domain specialist evaluation consisted of the same tasks that had been annotated during the error annotation (cf. section 4.4.12.1), 12 tasks per group. To facilitate the measurement of intra-annotator agreement, the agreement for the evaluators with themselves, 10% of the content (one task) was evaluated twice. Twenty-six evaluation task objects were uploaded for evaluation, 13 tasks per evaluator. The kappa coefficient (k) was used to calculate intra-annotator and pairwise inter-annotator agreement.

An example of the evaluation of a task is presented in Figure 4.4. The top part shows the German translations of the questions related to fluency and adequacy as

---

[34]https://github.com/cfedermann/Appraise

adapted from LDC (2005). Below that the reference translations of the active segment is displayed in bold with the preceding and following segment in regular font (for context).

The largest section of the interface displays five translation candidates for the active segment, containing four post-edited versions and the machine translated version. The origin of a segment (post-edited solution or machine translation) was not known to the evaluators. The translated labels for both scales (fluency and adequacy) are displayed above each translation candidate. As is evident from Figure 4.4, it was a monolingual evaluation task with all labels and descriptions presented in German.

FIGURE 4.4: Screenshot of evaluation task in Appraise

4.4.12.3   Community Evaluation

In order to ascertain how the post-edited content would be received by the end-users of the translated content, the importance of which has been stressed since the early days of MT evaluation (see e.g. van Slype 1979), an evaluation in the German Norton Community was devised. It takes the form of a bottom-up evaluation step (cf. Drugan

2013), which is then compared against the two top-down evaluation steps presented above (cf. section 4.4.12.1 and section 4.4.12.2), thus facilitating a comprehensive view of the quality perception of community post-editing.

The rating widget used in the German Norton Community was based on a previous project (cf. Mitchell and Roturier 2012), in which the comprehensibility of machine translated content (English-German) was rated by members of the German Norton Community in their forum. The widget is pictured on the left in Figure 4.5. It prompted the users of the Norton forum to rate whether the machine translated post displayed on the right was comprehensible or not.

FIGURE 4.5: Evaluation in previous MT evaluation project

Although rating required only two clicks, only 100 votes were collected in six months in a previous project (Mitchell and Roturier 2012). It was hoped to increase participation by making the content more accessible, i.e. by including the content to be evaluated on the front page and every other board page of the forum. In addition, the evaluation took place on segment rather than on post-level.

The content that was evaluated, fifty segments in total,[35] was randomly selected from the content that had been evaluated in the previous two evaluation steps, based on the pool of segments with an agreement of three or more evaluators for the fluency score

---

[35]For a list of these, see Appendix I

in the domain evaluation, which amounted to 668 raw MT output and post-edited segments. The set contained 15 machine translated sentences, 4 machine translated sentences that were submitted unchanged by one or more post-editors and 31 post-edited sentences. It was ensured that no sentence was included more than once in the set. A hundred-and-seventy-five of these were distinct source segments, of which ten segments were selected for each score. The reasoning behind selecting segments with scores ranging from 1 to 5 was to test the agreement between professional evaluators and community members on all levels. Fluency was evaluated, rather than adequacy, to keep the task as short as possible, as the language competences were unpredictable. Hence, the tasks were performed monolingually, i.e. with access to the raw MT output only. The community members[36] were presented a stripped-down fluency scale of that used in the domain specialist evaluation. As the evaluation was included on the front page of the community, a short description was included with the sample sentence and the scale, as can be seen in Figure 4.6. Guidelines were not included in order to keep the task as concise and simple as possible.

In order to ensure comparability to a certain degree between the three types of evaluation, the following was considered. Accuracy as employed in the error annotation was linked to adequacy in the domain specialist evaluation, as they both deal with the meaning transferred. Language in the error annotation was roughly represented by fluency in the domain specialist and the community evaluation. The error annotation differed from the domain specialist and the community evaluation in that it is based on the classification and quantification of errors, rather than assigning a score on aspects of translation quality.

---

[36]Here: any user visiting the German Norton Community regardless of whether they are registered or not during the evaluation period.

Rate the fluency of the following sentence*:

When a page opened, I noticed that the email had not been sent by a friend.

incomprehensible
disfluent
non-native
good
perfect
Vote

*This evaluation helps us to improve the quality of machine translation and to thus provide you with more content.

FIGURE 4.6: Community evaluation widget plus translation

The question for fluency, or "Sprachfluss" and "sprachliche Qualität" as used in German in the domain specialist evaluation, was shortened for the community evaluation, as presented in Figure 4.6. The two evaluation designs differed further in that the domain specialists were briefed on the purpose of the study and the evaluation in a face-to-face meeting and additional instructional emails, while this information was captured in one statement at the bottom of the rating widget for the community.[37] Community members only rated fluency and did not have access to a reference translation in contrast to the domain specialist evaluation.

The use of different typologies and the switching between languages prevented a consistent use of terminology throughout the thesis. To facilitate a comparison between the evaluation designs, nevertheless, the concept of 'Accuracy' is connected directly to 'Adequacy/Vollständigkeit' and 'Language' to 'Fluency/Sprachliche Qualität/Sprachfluss'.

---

[37]There was more information available for community members in a board dedicated to announcements.

4.5 Concluding Remarks

This chapter presented the methodology for both the pilot study and the full-scale study. Initially, the approach of a mixed methods research design was outlined, applied and justified as an approach for the current thesis. Subsequently, the design, procedure and observations made during the pilot study were described. Based on this, the revised research questions and corresponding hypotheses were derived and presented, followed by the operationalisation of the research questions and the variables. The experiment procedure was outlined briefly, followed by an in-depth discussion of the experiment design and the methods to measure the variables occurring in each step. The first involved the description and justification of the methods chosen to sample the participants and the content to be post-edited, the pre-editing and machine translation step and the operationalisation of the post-editor profile. The second presented the post-editing user interface and the data collection method. The last investigated the concept of post-editing quality, presented the final methods used to measure the same and the reasoning behind each choice, which facilitates the analysis of the data and the results are presented in the following chapters.

Chapter 5

Results I - Quantitative Analysis of Quality

5.1 Introduction

After outlining the methodology employed in this thesis, Chapter 5, Chapter 6 and Chapter 7 present a discussion of the data gathered in the post-editing experiment, the first focussing on PE quality from a quantitative perspective, the second on PE quality from a qualitative perspective and the third on the relation between the post-editor profile and quality. Chapter 5 investigates post-editing quality as measured by the error annotation, the domain specialist evaluation and the community evaluation. The results from each evaluation step are presented, followed by a discussion on quality in comparison to the raw machine translated output in the context of community post-editing, in answer to RQ1:

RQ1)*How does post-edited output produced by community post-editors compare to raw machine translated output?*

Figure 5.1 illustrates the data involved in answering RQ1. It consists of establishing counts of errors remaining in the MT and PE output, an evaluation performed by domain specialists and an evaluation performed by community members, i.e. end-users. As described in Chapter 4 section 4.4.12.1, 44% (1,738 segments) of the post-edited and the machine translated content were evaluated in the error annotation and the domain specialist evaluation steps. A subset of 50 segments were evaluated in the community evaluation (see Chapter 4 section 4.4.12.3).
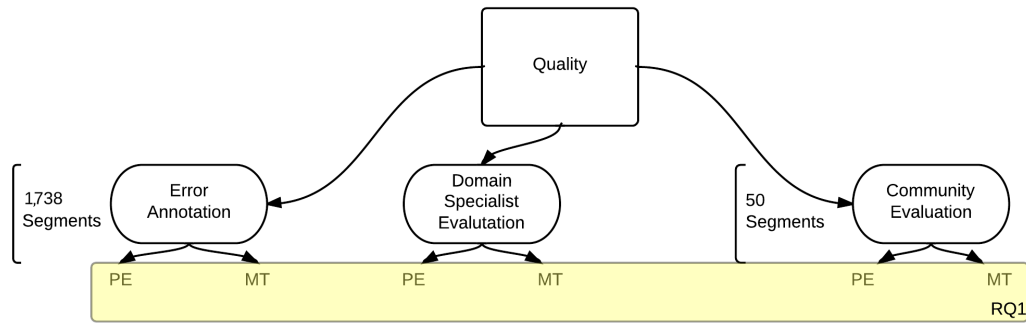
FIGURE 5.1: Exploring the concept of quality - RQ1

5.2   Error Annotation

This section focusses on the data obtained through the error annotation of the MT and the PE content. The error annotation involved identifying the errors remaining according to de Almeida's error categorisation (see Chapter 4 section 4.4.12.1), i.e. all errors present in both types of content. Errors were identified according to the definitions presented in de Almeida (2013, pp. 91-94). For the PE output, this was the case for errors regardless of whether they were present in the raw MT output or not. As

Number of errors	Total	in %	Accuracy	in %	Language	in %	Format	in %
0	**658**	**47**	**1,012**	**72**	**878**	**62**	**678**	**87**
1	331	23	250	18	316	22	99	13
2	176	12	88	6	132	9	5	1
3	110	8	28	2	55	4	1	< 1
4	56	4	16	1	24	2	2	< 1
5	34	2	11	1	6	< 1	0	0
6	28	2	6	< 1	3	< 1	0	0
7	10	< 1	0	0	0	0	0	0
8	5	< 1	1	< 1	0	0	0	0
9	4	< 1	2	< 1	0	0	0	0
10	2	< 1	0	0	0	0	0	0

TABLE 5.1: Error annotation - overview of segments with n number of errors

previously mentioned, the categories considered were Accuracy errors (additional information, missing information, untranslated information, mistranslated information), Language errors (grammar or spelling errors) and Format errors (missing or additional spaces). Format was considered separately, as incorrect spacing may not necessarily be regarded as degrading the post-edited content. This stands in contrast to incorrect punctuation. For example, misplaced commas may distort the flow of the sentence and missing commas may render a sentence very hard to read.

Table 5.1 presents a summary of the results of the error annotation by displaying the number and percentages of segments that contained 0 errors, in the Accuracy category and in the Language category, followed by the number and percentages of segments that contained 1 error in total, for Accuracy and for Language etc. It is evident that 47% of all post-edited segments contained no errors, while 72% contained no Accuracy errors and 62% no Language errors. While 47% may seem low, it needs to be noted that this includes data from all post-editors and all errors (also Format errors).

5.2.1 Focus on Accuracy and Language

In order to expand on the results presented above, the error categories Language and Accuracy are presented separately in the following table per post-editor. This enables an investigation of both concepts and the linking of Language to "fluency" and Accuracy to "adequacy" as used in the domain specialist evaluation in section 5.5. Table 5.2 displays the annotation results for all post-editors from both group A and B.[1] It displays the absolute number of errors remaining in the raw MT output and the PE content, followed by the ratio of PE errors over MT errors in percentage terms. The same format is used for errors in the Accuracy category, the Language and the Format categories. The final column contains the number of segments that were evaluated per post-editor.

	Overall			Accuracy			Language			Format			
	MT	PE	%	MT	PE	%	MT	PE	%	MT	PE	%	seg.
A1	234	163	70	112	87	78	117	61	52	5	15	300	81
A2	297	101	34	137	28	20	156	60	38	4	13	325	81
A3	280	82	29	114	31	27	161	48	30	5	3	60	76
A4	255	50	20	124	24	19	129	25	19	2	1	50	80
A5	270	59	22	126	16	13	141	34	84	3	9	300	82
A6	281	36	13	123	11	9	154	22	14	4	3	75	82
A7	252	45	18	120	22	18	128	20	16	4	3	75	75
A8	252	54	21	128	22	17	119	24	20	5	8	100	82
A9	265	54	20	126	19	15	135	30	22	4	5	125	85
AVG	*265*	*72*	*27*	*123*	*26*	*24*	*138*	*36*	*33*	*4*	*7*	*163*	*80*
B1	231	157	68	92	60	65	136	93	68	3	4	133	77
B2	221	68	31	101	16	16	115	50	43	5	2	40	74
B3	242	126	52	105	32	31	134	90	67	3	4	133	79
B4	232	53	23	93	19	20	137	27	14	6	2	33	77
B5	229	29	13	102	10	10	121	17	14	6	2	33	80
B6	264	61	23	113	17	15	144	39	27	7	5	71	75
B7	290	157	54	126	79	63	158	69	44	6	9	150	78
B8	281	197	70	121	113	93	155	75	48	5	9	180	76
B9	279	93	33	113	38	34	162	42	26	4	13	325	77
AVG	*252*	*105*	*41*	*107*	*43*	*39*	*140*	*55*	*40*	*5*	*5*	*157*	*77*

TABLE 5.2: Error annotation results

---

[1] These results are also presented and discussed in Mitchell, O'Brien and Roturier (2014).

On average, the raw MT output selected for evaluation per group A and group B contains a very similar number of errors (265 vs. 252). Within the groups, however, the number of errors in the raw MT output varies slightly, despite a random selection process of segments for each post-editor. In group B for both Accuracy and Language, the remaining number of errors are higher on average than for group A, which suggests that group A performed better overall than group B. The scope of errors remaining in the PE output for both groups ranges between 13% (A6, B5) and 70% (A1, B8). This illustrates that there is great variance in post-editing success across the post-editors. In contrast to group A with one post-editor (A1) producing PE output still containing $\geq 50\%$ errors, five post-editors in group B produced such output (B1, B3, B4, B7, B8).

Comparing the errors remaining in the PE output for Language and Accuracy, an average of 4.94% more errors corrected in the Language category can be observed, which is not statistically significant, however.

URLs contained the largest share of Format errors, i.e. additional and missing spaces, in the raw MT output (e.g. *Http: / / www.bleepingcomputer.com / http: / / www.geekstogo.com / Forum / http: / / www.cybertechhelp.com / forums / http: / / support.emsisoft.com / Forum / 6-help-my-pc-is-infected / http: / / forums.whatthetech.com /.* - Task A8, Segment 8.) Furthermore, post-editors introduced additional spaces in front of punctuation or at the beginning of sentences (e.g. *Echtzeit Scanner müssen alleine arbeiten.* in the post-edited content - Task B6, Segment 6). These observations, supported by the surprising increase in Format errors, may indicate that Format errors were not considered as high priority by the post-editors. This lack of attention to format may have been facilitated, for example, by the guidelines that promoted reusing "words, phrases or punctuation as they are", without, however, giving clear instructions on

formatting issues (cf. Chapter 4 section 4.4.10).

In summary, the number of errors remaining in the post-edited texts ranged from 13% to 70%, which shows that community post-editing can be feasible, depending on the post-editor. Results varied greatly across the post-editors with none of the participants producing output without errors.

De Almeida finds in her study that on average professional translators who post-edited texts in French and Brazilian Portuguese failed to correct 13.44% of all essential changes required and introduced 4.88% of new errors, adding up to a total of 18.32% errors remaining in the post-edited output (cf. de Almeida 2013, p. 186). This shows that the best performing post-editors (cf. Table 5.2: A6, A7, B5) of our study in the Norton Community were able to achieve results of the same or higher standard as the professional translators in de Almeida's study.

For French (en-fr), de Almeida found that 25% of the changes not implemented and newly introduced belonged to the Language category, while 71% belonged to the Accuracy category (de Almeida 2013, pp. 189-191).[2] For Brazilian Portuguese (en-bp), the numbers seem to be nearly reversed, as it is 27% of errors remaining, i.e. not corrected/newly introduced errors in the Accuracy category and 66% in the Language category (de Almeida 2013, pp. 192-194). Evidently, these results are very different to those presented here. A reason for this may be the suitability of the post-editors in general, professional translators versus volunteer community members or language-specific post-editing behaviour, depending on the output of the machine translation system and typical language specific mistakes, which is discussed in Rico and Ariano

---

[2]This is based on the interpretation of the numbers presented, summarised to include both Accuracy and Mistranslation errors, which are combined in the Accuracy category in the present study.

(2014), for example, focussing on the need for and the development of language specific post-editing guidelines for English-Spanish.

5.2.2   Focus on Sub-Categories of Errors

In order to investigate errors remaining and post-editing behaviour in more detail, the following section deals with the type of errors the post-editors corrected within the categories Language and Accuracy.
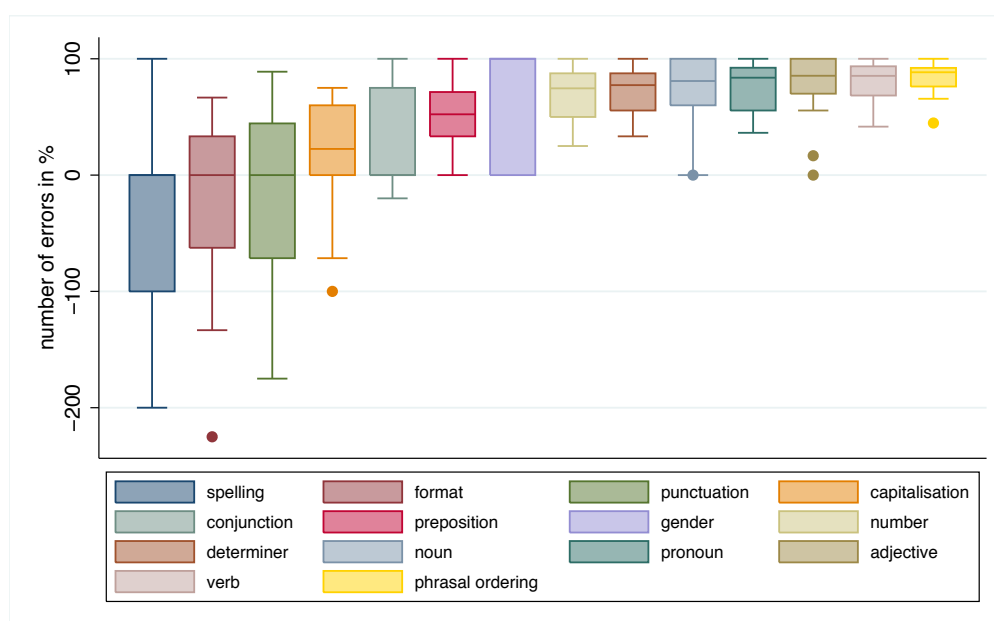
FIGURE 5.2: Overview: errors corrected/ introduced in % - Language category

Figure 5.2 shows the percentages of errors corrected from the sub-categories in the Language category across all post-editors.

It is evident that for over 50% of the data for spelling, format and punctuation, errors were introduced in these categories, rather than corrected. The median (horizontal line in the box) of corrected errors (in %) is 0. For capitalisation and conjunctions, there are still negative data points, with a positive trend as indicated by the raised median.

It is interesting to note that for format, punctuation and capitalisation, none of the post-editors corrected all errors present. This may be due to the fact that these three categories are subject to interpretation and preference, i.e. errors may not necessarily be perceived as errors, and errors in these categories may be tedious to rectify. While interpreting the data in these five categories, it needs to be considered that they had a low number of occurrences (average number per post-editor: 5 to 8).

As the median values indicate, the preferences for which errors to correct were close for numbers, determiners, nouns, pronouns, adjectives, verbs and phrasal ordering errors, with the highest number of occurrences in the category determiners, pronouns and phrasal ordering. Phrasal ordering was considered most important[3] and has the smallest spread of errors corrected ranging from 45% (outlier) to 100% and a median of 88%. Although the lower number of occurrences only allow for a tentative interpretation, the data collected suggests that verbs (85%), adjectives (85%) and nouns (81%) seem to have been considered high priority, as underlined by the small range of data points.

From Figure 5.3, it can be seen that mistranslations (83%) and extra information (82%) were corrected the most, followed by information missing (77%) with untranslated information being the least important category (69%).

In regards to Accuracy, the failure to correct 100% of missing information and mistranslation may have been due to the guidelines or due to a lack of training as a translator, e.g. insufficient SL skills. Presumably, correcting language errors requires less cognitive effort as it is performed monolingually, than correcting accuracy errors, which is performed bilingually. This gives rise to the question of why surface errors,

---

[3]Importance is based on the percentage of errors being corrected.

FIGURE 5.3: Overview: errors corrected/ introduced in % - Accuracy category

such as capitalisation or format errors were not corrected, as we assume they require even less cognitive effort than correcting grammatical errors. This may suggest that there is a certain threshold of cognitive effort below which the community members regard correcting errors as tedious and not beneficial, which results in errors below that threshold not being corrected.

In order to contextualise the results from the error annotation in the current body of research, they are compared to the French post-editing results gathered by de Almeida (2013) using the the same error annotation. Errors from the categories *mistranslation* (43%),[4] information missing (31%) and capitalisation (6%) are the categories that are most often not corrected. For the categories adjective, adverb, conjunction, gender, number, preposition and punctuation, between 95% and 100% of all errors present in the raw MT output were corrected. In de Almeida's data Accuracy errors are the type of errors that remain most often in the post-edited output, while there was no statistically significant difference in the data for this study. In terms of Language,

---

[4]Percentages of all errors not corrected were calculated from de Almeida's data by the present researcher.

there seems to be a parallel that correcting capitalisation errors is not of high priority, regardless of the setting.

It has to be noted that the distribution of errors in the raw MT output differs considerably between de Almeida's study and the present study, e.g. in de Almeida's study, errors of phrasal ordering account for one of the lowest error occurrence rates in the raw MT output, while it accounts for one of the highest occurrence rates in the present study, which is due to the language pairs under consideration. Hence, a comparison on a fine-grained level may not be beneficial.

5.3 Domain Specialist Evaluation

This section deals with the domain specialist evaluation to provide a second perspective on PE quality, using another top-down approach that is more high-level than the error annotation. The focus here is on domain specialists who are native speakers of the target language and (former) employees of Symantec either in the localisation or the technological support department. They rated adequacy based on their domain knowledge and fluency with reference to a gold-standard translation, i.e. monolingually (see Chapter 4 section 4.4.12.3). Due to the monolingual nature of the tasks, the German terms were presented to them: "Vollständigkeit" (adequacy) and "Sprachfluss" / "sprachliche Qualität" (fluency). Seven domain specialists (German native speakers) evaluated 905 (group A) or 865 (group B) segments. Of these, four were trained linguists as well as domain specialists. The evaluation took place over a period of two weeks. It was carried out using the open source evaluation system Appraise (Federmann 2010). The evaluators were not aware which segments were post-edited and which were machine translated. 44% of the post-edited content was selected and

combined randomly. Twelve tasks with one reference translation and four transla-
tion candidates provided by the post-editors plus one machine translated solution per
segment were created. Ten percent (group A: 17 segments, group B: 15 segments)
of this content was presented twice to the users, in order to measure intra-annotator
agreement. The evaluators completed 13 evaluation tasks.

Table 5.3 presents the overall results for the domain specialist evaluation. It displays
the share of segments per score for the post-edited content and the machine translated
output. Overall, it can be seen that post-editing was successful in terms of adequacy
with 67%, i.e. two thirds of all post-edited segments, rated as containing all content,
compared to roughly one third for all machine translated content. In terms of fluency,
39% of the post-edited segments were rated as "perfect" (5) and 34 % as "good" (4),
combined 73%. While it is crucial for the post-edited content to transfer all meaning
correctly (adequacy), it may not be crucial for the language (fluency) to be "perfect",
with a quality result of "good" possibly being sufficient. In this light, community
post-editing is considered as successfully increasing quality for both adequacy and
fluency compared to the raw MT output. It does not, however, reach the quality level
of professional translation, e.g. of that determined on average, 73%, in Sinaiko and
Brislin (1973). Despite the fact that the work by Sinaiko and Brislin seems outdated, it
makes for an excellent comparison, as they measure quality not only from three different
perspectives (using an error categorisation, knowledge testing and performance testing)
but as they design evaluation techniques that are very costly in terms of resources
and thus potentially more rigorous than current approaches to measuring translation
quality.

score	PE adequacy in %	MT adequacy in %	PE fluency in %	MT fluency in %
5	**67**	36	**39**	11
4	**20**	23	**34**	10
3	**6**	12	**16**	28
2	**4**	18	**6**	24
1	**2**	11	**5**	27

TABLE 5.3: Domain evaluation - PE segments (in bold) compared to MT segments per score, both in %

5.3.1 Domain Specialist Evaluation Per Post-Editor

This section aims at explaining the origin of the numbers presented in Table 5.3, focussing on the performance per post-editor. Figures 5.4 and 5.5 show the number of segments for each score (e.g. 1 = "incomprehensible" for fluency and "none" for adequacy) per post-editor for both adequacy and fluency with the post-editors on the x-axis and the evaluation scores on the y-axis. This includes the scores for the raw MT output (MT A and MT B in Figure 5.4 and 5.5). The results are ordered by the share of segments with the best score ("5") and show the ranking (increasing) of the post-editors and the machine translated content based on the share of segments rated as "5".
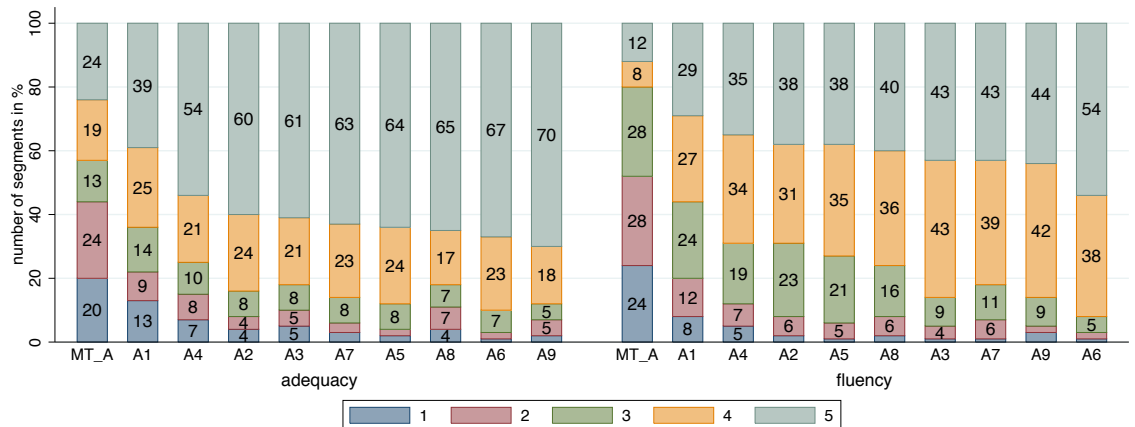
FIGURE 5.4: Number of segments per score and post-editor for adequacy and fluency in %, group A
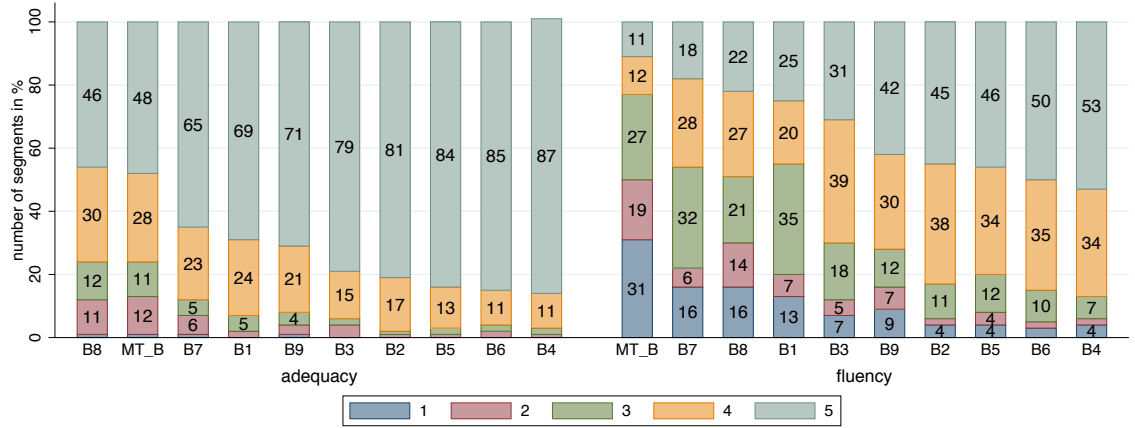
FIGURE 5.5: Number of segments per score and post-editor for adequacy and fluency in %, group B

It is evident from Figure 5.4 that all post-editors were able to increase the share of segments with a score of 5 for both adequacy and fluency. Apart from A1, the share of segments with all meaning preserved (score "5" adequacy) was increased to 50% or higher with 70% of all segments reaching a adequacy score of "5" for the best performing post-editor. For group B, cf. Figure 5.5, there was one post-editor who scored slightly lower for adequacy than the raw MT output was scored. Overall, adequacy was scored considerably higher from the onset for group B; 24% of segments with a score of "5" for the raw MT output of group A and 46% of segments with a score of "5" for the raw MT output of group B. The possibility of the MT output naturally containing fewer errors for group B than for group A can be ruled out based on the results of the error annotation, which indicated an equal number of errors for both groups in the MT output. On the other end of the spectrum, the best performing post-editor for group A achieved a share of 70% with a score of "5", while it is considerably more for group B with 87% of segments with a score of "5". Again, it is useful to remember that even professional translators do not achieve 100% on average in quality evaluation processes.

Figure 5.4 shows that, for group A, considerably fewer segments reached a score of "5" in fluency before and after post-editing compared to the results for adequacy. This discrepancy can also be observed for group B, as presented in Figure 5.5. None of the post-editors reached a share of above 55% of segments with a score of "5". This may be due to a number of reasons. Firstly, the guidelines specified to reuse as many structures from the raw MT output as possible. Secondly, the focus may have been on the content rather than the language while post-editing. Thirdly, editing the language may have been considered too tedious in some cases. Of course, language can never be accurately distinguished from content. To facilitate an analysis on a macro-level, it is assumed here that, on average, a distinction could and was made between language and content errors.

This shows that the best performing post-editors are able to achieve quality levels in terms of adequacy approaching that of professional translators as determined by similar studies, but not in terms of fluency.

These results somewhat contradict the results from the Error Annotation presented above, in which a comparable number of errors remained in both Language and Accuracy. A reason for this could be that it is cognitively less demanding to evaluate 'fluency" than "adequacy", as the first is evaluated with reference to the TT only and the latter with reference to both the TT and a reference translation.

5.3.2 Domain Specialist Evaluation Per Evaluator

This section investigates the homogeneity of the two groups of evaluators in order to determine where the evaluator variance stems from.

	MT adequacy	PE adequacy	difference	MT fluency	PE fluency	difference
group A	2.98	4.28	1.30	2.50	4.05	1.54
group B	4.13	4.64	0.52	2.59	3.86	1.27

TABLE 5.4: Comparison between raw MT output and PE results, average results

Table 5.4 displays the results from the evaluation of the raw MT data compared to the evaluation of the corresponding PE content for both group A and B. The first column specifies the average scores by evaluator compared to the output produced by each post-editor.[5] This includes all data obtained through the domain specialist evaluation, i.e. 44 % of the post-edited content plus the repeated segments for measuring intra-annotator agreement. The second column contains the average of the adequacy scores for the raw MT output per evaluator. The third column presents the average adequacy scores for the PE produced for each evaluator-post-editor pair. The fourth column shows the average difference between the two preceding values. Column five contains the average fluency scores for the raw MT output and column six for the post-edited output, which is followed by the average difference.

For group A, it was the case for all post-editor evaluator pairs (individually and on average) that the ratings improved compared to the raw MT output with (high) statistical significance, which is an indicator of very consistent scoring.

For group B, adequacy and fluency increased for all post-editor evaluator pairs (individually and on average), however, to a smaller extent than for group A with predominantly statistical significance. For one post-editor, the scores were only marginally higher than the MT output, an improvement that was without statistical significance. The individual scores that this analysis and the average scores in Table 5.4 are based on are displayed in Appendix J, Table J.1.

---

[5]The full tables can be found in Appendix J.

Comparing both groups, it is apparent that the approach to the evaluation seemed to have been different for the evaluators of group A and group B in terms of adequacy. Group A rated adequacy of the MT output on average a lot lower (= 2.98) than group B (= 4.13). That is why there is only a small margin for improvement for the post-edited output in group B. However, the results in Table 5.4 suggest that adequacy was rated very high in general for the raw MT output, while the error annotation suggests that there was a comparable number of errors in the raw MT output for group A and group B.

To summarise, the domain specialist evaluation revealed the following: in total, 67% of segments evaluated achieved a score of "all" ("5") in terms of adequacy and 73% of segments achieved a score of "perfect" ("5") or "good" ("4") in terms of fluency. Furthermore, there was a large variance in adequacy scores across the evaluators. In group A, the share of segments that was rated as "all" ranged from 24% - 70%, while it ranged from 46% - 87% for group B. Between 12% - 55% of segments achieved a score of "perfect" for group A, while it was 11% - 53% for the post-editors of group B. The evaluation confirmed that all post-editors were able to improve the machine translated output for both adequacy and fluency compared to the MT output with statistical significance (bar one post-editor in group B). It is noteworthy that the evaluators for group B appeared to take a different approach to scoring adequacy, which resulted in higher adequacy scores for the raw MT output and the post-edited content.

5.3.3 Intra- and Inter-Annotator Agreement

As touched upon above, there was some variance in evaluator judgement across the groups. In order to assess the homogeneity/heterogeneity within the groups of evaluators and the consistency for each individual evaluator in more detail, inter- and intra- annotator agreement was calculated for the domain specialist evaluation. Cohen's Kappa was employed, a standard measure to calculate annotator agreement in evaluation tasks (see for example WMT 2012). As the evaluation involved rating segments on a categorical, rather than a numerical scale, the definition for weighted kappa was used for calculating intra-annotator agreement, i.e. the rate that they agree with themselves, and the inter-annotator agreement, i.e. the rate that they agree with other evaluators.

$$(1) \ k = \frac{P_o - P_e}{1 - P_e}$$

$$(2) \ w_{ij}(i = 1, ..., k), \ (j = 1, ..., k) \text{ and } w_{ij} = 1 - \frac{|i-j|}{(k-1)}$$

(1): with P_o as the observed proportion of agreement (also called actual proportion of agreement P(A), cf. WMT 2010) and P_e as the expected proportion of agreement.

(2): with i containing the ratings of the first evaluator and j the ratings for the second evaluator (or the same evaluator for intra-annotator agreement) and w_{ij}, hence, defining the weights for agreement and disagreement (StataCorp 2014). A kappa coefficient value of 0 is returned when the agreement observed is expected to have occurred by chance. A value of 1 indicates perfect agreement. For intermediate values, the interpretation devised by Landis and Koch (Landis and Koch 1977, p. 165) is applied.

Table 5.5 shows the results for intra-annotator agreement with particularly low agreement in grey.

Evaluator	P_o in %	P_e in %	k	P_o in %	P_e in %	k
Group A		*adequacy*			*fluency*	
E1	90.00	67.22	0.70	89.12	68.44	0.66
E2	82.65	64.25	0.51	85.88	66.72	0.58
E3	86.47	63.98	0.62	87.65	61.50	0.68
mean			0.61			0.64
Group B						
E4	36.00	37.67	-0.03	87.67	63.53	0.66
E5	95.33	80.98	0.75	92.67	70.42	0.75
E6	95.56	84.52	0.71	88.67	72.92	0.58
E7	78.00	73.11	0.18	88.67	74.95	0.55
mean			0.41			0.64

TABLE 5.5: Intra-annotator agreement for the domain specialist evaluators

Table 5.6 displays pairwise inter-annotator agreement, split into group A and B with particularly low agreement in grey.

Evaluators	P_o in %	P_e in %	k	P_o in %	P_e in %	k
Group A		*adequacy*			*fluency*	
E1 - E2	82.90	67.67	0.47	85.69	69.21	0.54
E1 - E3	82.90	68.46	0.46	82.82	65.79	0.50
E2 - E3	80.61	66.05	0.43	81.66	64.58	0.48
mean			0.45			0.51
Group B						
E4 - E5	82.37	79.91	0.12	75.38	60.27	0.38
E4 - E6	87.63	83.97	0.23	77.75	59.95	0.44
E4 - E7	85.47	82.38	0.18	77.14	60.49	0.42
E5 - E6	85.43	76.78	0.37	85.14	66.45	0.56
E5 - E7	84.94	78.91	0.29	81.30	65.46	0.46
E6 - E7	87.14	82.45	0.27	81.18	64.52	0.47
mean			0.24			0.46

TABLE 5.6: Inter-annotator agreement for the domain specialist evaluators per group

According to the interpretation of scores by Landis and Koch (1977), intra-annotator agreement ranges from poor to substantial for adequacy and from moderate to substantial for fluency. As expected, intra-annotator agreement is higher than inter-annotator agreement, as it is substantial for group A for both adequacy and fluency and moderate for adequacy for group B and substantial for fluency for group B. Inter-annotator agreement is higher for fluency than for adequacy in both groups, whereas intra-annotator

agreement is higher for adequacy in group A than for fluency, while it is mixed for group B. This is noteworthy, as it shows that there does not seem to be a pattern for intra-annotator agreement, while the agreement scores for inter-annotator agreement seem to suggest that it is "easier" to achieve high agreement on fluency scores. This may be based on the fact that rating a segment's fluency, mainly based on language, is easier than rating a segment's adequacy, mainly based on content, while referring to a reference segment.

Both the intra-annotator agreement for adequacy and the inter-annotator agreement for adequacy is particularly low for E4 and E7. The low intra-annotator agreement for adequacy for those two evaluators is likely to be a reason for the low overall inter-annotator agreement for adequacy - the highest agreement in that category is for evaluators E5 and E6, which is, however, still below the average for group A. In summary, this section supported the findings of the error annotation that the post-editors improved the raw MT output, however, with a slightly different emphasis in regards to Accuracy/adequacy and Language/fluency. Furthermore, an investigation into evaluator agreement revealed that most evaluators were consistent within themselves but that there was great variance between different evaluators of the same group and across the groups, which is a common observation in similar studies in the field.

5.4   Community Evaluation

As discussed in detail in Chapter 4, a community evaluation was devised to gather feedback from the end-users of the post-edited content.

The total number of words in the fifty sentences selected for the community evaluation was 658. The total number of errors identified in these fifty sentences during the error annotation phase was 136, with 61 (44.85%) errors in the Accuracy category, 72 (52.94%) errors in the Language category, and 2 (1.47%) in the Format category. The community evaluation took place over four weeks, with all fifty post-edited sentences being available for evaluation in the first two weeks. During the second two weeks sentences 21-50 were displayed only, in order to increase the number of ratings for these sentences. There were 1,470 ratings with eight as the lowest number and 110 as the highest number of ratings received per segment. The median number of ratings per segment was 21.5. The community evaluation was performed by 171 discrete users, 143 of whom were unregistered and 28 of whom were registered in the community. The average number of ratings per user session (in one sitting) was eight ratings. The fifty segments received 29 ratings on average, with the highest number of ratings being 110 and the lowest number of ratings being 8. An overview of the results is displayed in Table 5.7. The first two columns show the absolute number of ratings received in the community evaluation as "incomprehensible", "disfluent" etc., followed by the percentages.

	segments rated as	%
incomprehensible	457	31.1
disfluent	208	14.1
non-native	387	26.3
good	270	18.4
perfect	148	10.1

TABLE 5.7: Community evaluation results

The largest share of ratings (31%) was "incomprehensible" ("1") followed by "non-native" ("3") with (26%). The least chosen category was "perfect" ("5", 10%), the best category followed by "disfluent" ("2", 14%). The results were not expected to

have an equal share of 20% per category. Despite this, these numbers may show a trend towards a stricter evaluation by the community.

5.4.1 Inter-Annotator Agreement

For the community evaluation, due to the nature of the ratings, a varying number of evaluators (> 2) with a varying number of ratings per segment (> 2), agreement was calculated using the basic definition for kappa (cf. Section 5.3), rather than using weighted kappa. Furthermore, agreement was not calculated pairwise but on a bulk level. The mean of kappa was calculated as follows:

$$\overline{k} = \frac{\sum_j \overline{p}_j \overline{q}_j \hat{k}_j}{\sum_j \overline{p}_j \overline{q}_j}$$

The inter-annotator agreement (mean) for the community is fair ($\overline{k}$=2.1). Comparing this to the results for WMT 2011 ($\overline{k}$=3.8), which were based on rankings, rather than ratings on categorical scales, we see that inter-annotator agreement is rather low here. For the WMT task, ratings were obtained from participants of the shared tasks only. The pool of evaluators was then diversified by opening it up to all members of Amazon Mechanical Turk for WMT 2012, which is comparable to the pool of evaluators for the community evaluation, as it was open to anyone who visited the German Norton forum. Tapping into a more diverse pool of evaluators, consisting of users of AMT, Callison-Burch et al. (2012) recorded a lower inter-annotator agreement ($\overline{k}$=2.8) than for their 2011 evaluation tasks, which is comparable to the agreement measured here.

5.5   Discussion of RQ1

After presenting the results of the three different types of evaluation and thus three different perspectives on quality, this chapter will combine and discuss the results in order to answer RQ1. This thesis investigates whether it is feasible for community members of the Norton Community to post-edit UGC that has been machine translated from English into German. The baseline of feasibility here is identified as being at least of higher quality than raw machine translated output. In an attempt to establish how quality may be defined in an online community context with a different primary purpose than that of translation/post-editing, it was measured in three different ways, as outlined in Chapter 4. The following research question is answered and the corresponding hypotheses are thus tested.

RQ1) How does post-edited content produced by community post-editors compare to raw machine translated output?

H1.1$_0$) Community post-editing does not produce content that is of superior quality to raw machine translated output.

H1.1$_1$) Community post-editing produces content that is of superior quality to raw machine translated output.

H1.2$_0$) There are no correlations between the results of the error annotation, the domain specialist evaluation and the community evaluation.

H1.2$_1$) There are correlations between the results of the error annotation, the domain specialist evaluation and the community evaluation.

We hypothesise firstly, that the output produced by community post-editors here is going to be of better quality than the raw machine translated output. Secondly, we hypothesise that there are correlations between the different types of evaluation. Thirdly, we hypothesise that there is some overall correlation between the three quality evaluation approaches at a macro level. We hypothesise further that the evaluation steps chosen differ conceptually from each other, suggesting that 'quality' has many different components, which when measured individually yield different results at a micro-level.

In order to confirm or dismiss $H1.1_0$, it is helpful to revisit the results gained individually in the three evaluation steps. As presented in section 5.2, the linguistic evaluation showed that post-editing by members of the online community was feasible because there was a demonstrative improvement compared to the raw MT output, but that the post-editing performance varied across post-editors, ranging from 13% of errors to 70% of errors remaining in the post-edited content, with an average of 27% (group A) and 41% (group B) remaining.

The domain specialist evaluation showed that on average 53%/69% (group A/B) of the content received the highest score in terms of adequacy, whereas 35/32% (group A/B) received the highest score in terms of fluency. Furthermore, the domain specialist evaluation indicated that 67 % of all segments were rated as "5" (containing "all meaning") for adequacy, whereas it was 62% for the top two categories ("perfect" and "good") fluency. A comparison of the domain specialist evaluation of the raw machine translated output and the post-edited content revealed that there was a statistically significant improvement across all post-editors (except for one post-editor) for both fluency and adequacy.

Based on the linguistic and domain specialist evaluation data, $H1.1_0$ was rejected. We

can conclude from this that community post-editors can perform a reasonably good level of post-editing in relation to improving adequacy, but that fluency improvements are less apparent. It is difficult to ascertain why adequacy improved more than fluency, but it may be attributed to the guidelines that were available (optional) to the post-editors or the fact that these community members are not linguists for the most part and so style and fluency may be of less priority to them than to the domain specialist evaluators with advanced language skills, some of whom work in the field of localisation. Furthermore, in the context of user-generated content, people are used to unstructured, unpunctuated and ungrammatical content, which may have lowered the importance of fluency.

In a second step, it is worthwhile comparing the three evaluation approaches to confirm or dismiss $H1.2_0$, i.e. whether there are correlations present between the three different evaluation types, or not.

Table 5.8 displays the correlation for the total number of errors, accuracy errors and language errors present in the post-edited content with the domain specialist and the community evaluation with * as statistically significant ($p = \leq 0.05$) and ** as highly statistically significant ($p = \leq 0.01$). Fluency scores are considered here only, as it facilitates a comparison between the domain specialist and the community evaluation.

	domain specialist	community
errors after PE	-0.65**	-0.61**
Accuracy (A) errors after PE	-0.74**	-0.75**
Language (L) errors after PE	-0.67**	-0.66**
edited	0.58**	0.61**

TABLE 5.8: Correlation between the error annotation and the domain specialist and community evaluation (set of 50 segments)

As displayed in Table 5.8, there is a moderate negative correlation between the total
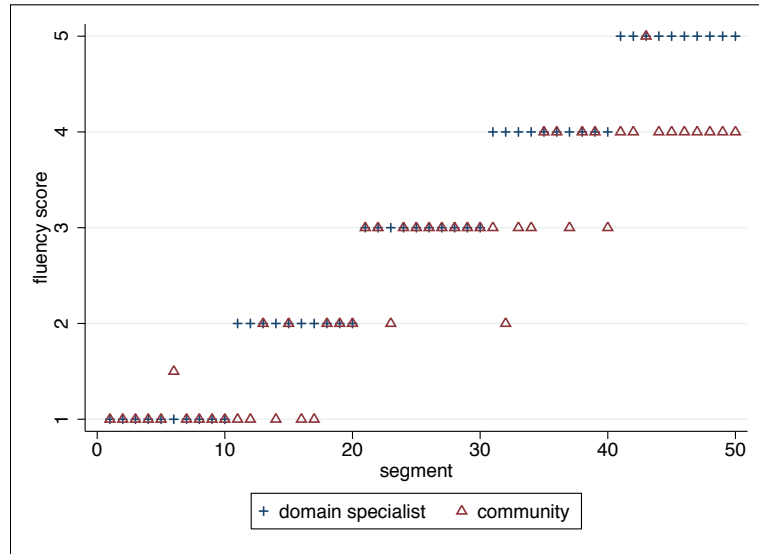
FIGURE 5.6: Domain specialist and community evaluation ordered by score

number of errors present and the domain specialist/community evaluation, as well as accuracy errors and language errors in the evaluated output. These relationships are statistically significant. There is a moderate significant correlation between the editing of the raw MT output and the evaluation score, i.e. if the segment was edited, the fluency score was likely to increase. To conclude, the community evaluators and domain expert evaluators behaved in a similar pattern in that they rated segments lower, which contained more errors for either one of the categories of Accuracy and Language or both of them. Similarly, if the segment was edited, both types of evaluators rated the segment higher.

5.5.1 Domain Expert vs. Community Evaluation

Figure 5.6 compares the results from the domain expert evaluation to that of the community evaluation in more detail. It displays the domain expert ratings and the community ratings for each segment, ordered by the scores of the domain expert evaluation.

As can be seen, there were a total of 28 (56%) equal ratings (triangle containing plus sign) between the domain expert evaluation and the community evaluation (median), meaning that just over half the time, the two sets of evaluators agreed. In 21 cases (42%), the community rated the content lower on the scale compared to the domain expert evaluation (triangle is situated below plus sign). This implies that the community evaluators were inclined to rate segments more negatively than the domain experts. In one case (2%, segment 6), the community rated the content higher on the scale than the domain expert evaluators (triangle is situated above plus sign). Only in one case does the community evaluation deviate more than one point on the scale from the domain expert evaluation (segment 32). The community evaluation deviates 0.05 points at least and 0.9 points at most, with an average of 0.45 points. Agreement was lowest for the segments that were rated as "5" and highest for the segments that were rated as "1" or "3" by the domain expert evaluators.

Although the results of the domain specialist evaluation and the community evaluation are strongly correlated (Pearson's Correlation: 0.92), it is interesting to establish the delta that does exist between the two evaluation processes and whether this difference occurred at random or not. A two-tailed t-test confirmed that there is a statistical significance of 0.43 points ($p < 0.001$) between the domain specialist evaluation and the community evaluation. This means that if the professional evaluation scored a segment high on the scale provided, so did the community evaluators, albeit with a slightly lower score. Hence, the community evaluators were slightly stricter in their evaluation as measured by the fluency scores.

5.6 Concluding Remarks

This chapter presented the quantitative data collected in the error annotation, the domain specialist evaluation and the community evaluation in answer to RQ1. The data presented above confirms that community post-editors were able to produce post-edited content that is of superior quality to raw MT output $(H1.1_1)$. Furthermore, there is a strong correlation (Pearson's Correlation 0.92) between the domain specialist and the community evaluation. There were correlations present that indicated a general relationship between an increasing number of errors and a lower fluency score in the domain specialist and the community evaluation, which were not statistically significant. This suggests that post-edited quality as measured by domain specialists and the community according to the concept of adequacy cannot be represented by the number of errors, as obtained through the error annotation. $H1.2_0$) can be rejected for the domain specialist evaluation but not for the error annotation. Despite the fact that the domain specialist and the community evaluation are strongly correlated, a statistically significant difference of 0.43 points ($p < 0.001$) was recorded. The next chapter will now discuss PE quality from a qualitative perspective.

Chapter 6

Results II - Qualitative Analysis of Quality

6.1   Introduction

After analysing post-editing quality from a quantitative perspective, this chapter investigates the post-editing quality on a more granular level. It aims at identifying when post-editing is successful, when it is unsuccessful (cf. Koponen 2011) and how this is manifested in edits. It analyses a subset of post-edited segments on both ends of the quality spectrum and seeks to answer the following question:

RQ2) *What are segment characteristics and post-editing patterns that render post-editing successful/unsuccessful?*

6.2   Segment Characteristics

This section deals with characteristics of all segments, i.e. the length of a segment (in words) before and after post-editing and the number of errors and error types per segment, and how these relate to the quality. Table 6.1 displays Spearman's correlation between these characteristics and the scores obtained through the domain specialist evaluation for adequacy and fluency for all segments evaluated (1,416 segments) with * as statistically significant and ** as highly statistically significant.

	adequacy	fluency
MT length	-0.11**	-0.33**
PE length	-0.11**	-0.31**
total errors	-0.44**	-0.62**
Accuracy	-0.43**	-0.50**
Language	-0.27**	-0.46**
Format	-0.04*	-0.14**

TABLE 6.1: Correlations between segment characteristics and domain specialist evaluation

According to Table 6.1, segment length appeared to influence fluency slightly. Accuracy errors appeared to influence adequacy and fluency equally, while Language errors were more likely to influence fluency scores. Format errors did not seem to influence either adequacy or fluency.

Next, it is of interest to categorise the segments according to their quality scores, into low-scoring and high-scoring segments to identify any unsuccessful and successful post-editing behaviour.

High-scoring segments were defined as having an average value of 5 for adequacy *and* for fluency obtained in the domain specialist evaluation. The group of low scoring segments was defined as having obtained an average value of ≤ 3 for adequacy *or* fluency in the

domain specialist evaluation. The cut-off point for high-scoring segments is the highest quality possible, as the community evaluation showed that they are very strict in their evaluation and in order to focus on excellent post-editing (as identified by domain specialists) rather than mediocre post-editing results. The definition of low-scoring segments is broader, to capture the range of unhelpful edits/errors. In the high-quality group, 99 segments were returned with 48 distinct segments (see Appendix K for a list of the segments), whereas it was 310 with 276 distinct segments in the category of low-quality segments (see Appendix L).

Before exploring post-editing examples, this section revisits the characteristics presented in Table 6.1 in the light of high-scoring and low-scoring segments. Table 6.2 compares the length of segments in words for the raw MT segments and the PE segments. It focusses on the total number of errors per segments and the number of accuracy, language and format errors. As the segment length differs considerably between the low scoring and high-scoring segments, the table also displays the average number of errors per 10 words.

	MT length	PE length	total errors	Accuracy	Language	Format	errors per 10 words
low quality							
average	14	13.6	2.9	1.5	1.1	0.1	2.6
maximum	49	47	10	9	6	5	40
minimum	2	1	0	0	0	0	0
high quality							
average	10	10.1	0.2	0.02	0.1	0.04	0.2
maximum	27	27	2	1	2	1	2.9
minimum	1	1	0	0	0	0	0

TABLE 6.2: High-scoring vs. low-scoring segments (per segment)

Low-scoring post-edited segments had an average length of 14 words, while the average length was 10 words for high-scoring segments before and after post-editing. Shorter

segments were rated better on average. There was a difference of 20 words for the maximum length of segments, while the minimum segment length was 1 word for both. There was no difference between the number of accuracy and language errors on average, which was the case for both low-scoring and high-scoring segments. The average number of errors per 10 words was 2.6 for low-scoring segments and 0.2 words for high-scoring segments. The maximum number of errors per 10 words (low-scoring) in the sample obtained, was an exception with a value of 40, as the segment in question consisted of 1 word but contained 4 accuracy errors (information missing).

6.3 When Does Post-Editing Fail?

Table 6.3 displays the number of low-scoring segments that contain one or more errors from any one error category. It is evident that the error types belonging to the main category Accuracy (except for untranslated words) appeared to be most disruptive to fluency and adequacy, according to the domain specialist evaluation. Errors of the type phrasal ordering seemed to influence the quality negatively. The number of segments with more than one error in one error category was $\leq 8\%$ (i.e. max. 25 segments per category).

error type	frequency	percent	category
gender	3	1	Language
conjunction	10	3	Language
spelling	11	4	Language
adjective	14	5	Language
untranslated	18	6	Accuracy
capitalisation	18	6	Language
number	17	6	Language
pronoun	17	6	Language
noun	20	7	Language
verb	31	10	Language
preposition	33	11	Language
punctuation	33	11	Language
determiner	40	13	Language
extra information	55	18	Accuracy
phrasal ordering	83	27	Language
information missing	94	30	Accuracy
mistranslated	134	43	Accuracy

TABLE 6.3: Number of low-scoring segments containing ≥ 1 error(s) of each error type

Some patterns that caused low quality scores in the domain specialist evaluation for both adequacy and fluency can be categorised as follows:

1) Source text issues:
- poor sentence structure
- ellipses (e.g. subject lines)
- needed explicitation
- use of colloquial language

2) Machine translation issues:
- mistranslations
- unidiomatic translations

3) Post-editing process issues:
- mistranslations (based on lack of SL skills)
- carelessness (obvious errors introduced that could have been easily fixed by the post-editor, such as an extra letter)
- unidiomatic phrases (e.g. literal translations of terms that can be understood but are not correct)
- paraphrasing (e.g. deviating 'too much' from the meaning expressed in the ST)
- summarising (i.e. leaving out vital information)
- no equivalent concept in German

It is not surprising that the source text created some of the difficulties that the community post-editors struggled with, since it was user-generated content. The pre-editing step that consisted of three rules, "Sentence too long", "Sentence ending punctuation" and "Spelling error" (cf. Chapter 4, section 4.4.6), was not sufficient to address all issues in the source text. A prevalent problem that the post-editors faced was poor MT output, due to poor sentence structure in the source text. This included ellipses, such as subject lines, and sentences that lacked explicitation, i.e. leaving out

optional components, such as "that" or the subject of the sentence. The average adequacy score was 3.1 for the raw MT output and 3.5 for the post-edited output. Hence, even in the low scoring category, the post-editors were able to improve adequacy by 0.4 points with statistical significance. The average fluency for the raw MT output was 2.0, which was improved by 0.6 points with statistical significance for the low-scoring category.

6.3.1 ST and MT Issues

In the present section, examples of ST issues will be investigated. The first issue to be examined is that of ellipses. Fourteen distinct subject lines[1] (which constitute 58% of all 24 distinct subject lines) and 29 subject lines in total fell into the low-scoring category. Below, two examples of elliptical sentence structures are displayed as source, machine translation (plus literal gloss),[2] post-editing solution (plus literal gloss), the adequacy and fluency scores for MT and PE output, and the post-editing distance comparing the MT output with the post-edited output as measured by TER. This format is used for displaying all post-editing examples in all tables throughout this section.

The first example in Table 6.4 is a typical subject line in the Norton Community. It contains the keywords to quickly describe the problem, without forming a complete and coherent sentence, which constitutes a problem for the MT engine. The post-editor was able to fix some errors the MT system had produced. However, they transferred the ST sentence structure, which is not idiomatic in German and was penalised by the

---

[1]While there were no duplicates in the post-edited sentences, some of them originated from the same source segment.

[2]Literal gloss is abbreviated to "gloss" in all tables throughout this chapter.

	segment	adequacy	fluency	TER
ellipsis 1				
ST	Copy to clip board not working for red threats detected screen			
MT	Kopieren, clip Board für Red erkannte Bedrohungen Bildschirm"	1.33	1.67	
gloss	Copy, clip board for red detected threats screen"			
PE	In Zwischenablage kopieren fehlerhaft für Bildschirm "Red Threats erkannt"	2	2.67	73%
gloss	Copying to clip board malfunctioning for "red threats detected" screen			
ellipsis 2				
ST	When I do ctrl ALT delete and click start task manager and then go under processes.			
MT	Wenn ich ctrl ALT löschen und klicken Sie auf Start der Aufgabe Manager und fahren Sie unter Prozesse.	1.33	1.33	
gloss	When I ctrl ALT delete and you click on Start the task manager and you drive under processes.			
PE	Mit ctrl Alt löschen und dann bei dem Aufgaben manager auf Start klicken und somit die Prozesse zu beenden.	3	2	70%
gloss	With ctrl ALT delete and then click Start in the task manager and thus end the processes.			

TABLE 6.4: PE examples for elliptical sentence structures

domain specialists. Presumably, the post-editor would not have been able to achieve top scores without diverging from the ST and forming a complete sentence, when the source sentence was incomplete already. This raises the question of whether the source text should also be rated when looking at post-editing quality and whether post-editing guidelines should specify how to proceed with ellipses. The second example shows another incomplete sentence, in which the MT system struggles with translating the verbs, which the post-editor corrects. The post-editor introduces errors of meaning, however, and retains the incomplete sentence, both of which are reflected in lowered scores.

The example in Table 6.5 shows a typical sentence structure occurring in the Norton

	segment	adequacy	fluency	TER
sentence structure				
ST	This is because the two fought over what to do with whatever they found that was suspect.			
MT	Dies ist, weil die beiden gekämpft, was zu tun ist, mit dem, was sie gefunden, die verdächtig war.	3.5	1.5	
gloss	This is because the two fight, what is to be done, with this, which they find, whom was suspect.			
PE1	Dies deshalb, weil die beiden Programme miteinander kämpfen, was zu tun ist bei verdächtigen Aktionen.	4.5	2.25	83%
gloss	This is because the two programs fought over what is to be done in case of suspicious activity.			
PE2	Dies passiert, weil die beiden kämpfen, das zu tun und damit das, was sie gefunden haben, ist verdächtig geworden.	4.5	2	42%
gloss	This happens because the two fight to do this and so that which they found has become suspicious.			

TABLE 6.5: PE examples for complex sentence structures

Communities. It does not contain commas and complexity is increased by multiple clauses. The MT system inserts commas and fails to convey the meaning of the sentence due to translating verbs incorrectly and leaving out relations between the different clauses. PE1 manages to eliminate the complexity introduced by the MT engine by rephrasing "dem, was sie gefunden, die verdächtig war" as "verdächtigen Aktionen" and thus removing a sub-clause. PE1 explicitates[3] the subject of the sentence "die Programme" but mistranslates the second half of the sentence. PE2 keeps the sentence structure of the source and the MT output and also mistranslates the meaning of the second half of the sentence. It is evident here that the domain specialists did not consider the content as problematic in the evaluation but rather the structure and the fluency of the sentence.

The first example of Table 6.6 shows that the MT system struggled with "shown

---

[3]Explicitation refers to "the spelling out in a target text of information which is only implicit in a source text" Olohan and Baker (2000, p. 142).

	segment	adequacy	fluency	TER
explicitation 1				
ST	Then select the 'Export Results' link shown below to export the scan results to a text file:			
MT	Wählen Sie den Link "Exportergebnisse", um die Prüfung unten angezeigten Ergebnisse in eine Textdatei:	3.33	2.33	
gloss	Select the "Export results" link to the scan results shown below to a text file:			
PE	Wählen Sie den Link "Exportergebnisse", um die Prüfung der unten angezeigten Ergebnisse in einer Textdatei:	2.67	2.67	11%
gloss	Select the "Export results" link to the scan of the results shown below to a text file:			
explicitation 2				
ST	All you need to do is enter your new Activation key in to your current subscription before it expires.			
MT	Alles, was Sie tun müssen, ist, geben Sie Ihren neuen Aktivierungsschlüssel in Ihrem aktuellen Abonnement, bevor es abläuft.	4.5	3.25	
gloss	All you need to do is you type your new activation key into your current subscription, before it expires.			
PE	Alles, was Sie tun müssen, geben Sie Ihren neuen Aktivierungsschlüssel in Ihr aktuelles Abonnement, bevor es abläuft.	4	3	19%
gloss	All you need to do is, you type in your new activation key into your current subscription, before it expires.			

TABLE 6.6: PE examples for STs lacking explicitation

below" (an ellipses of "that is shown below"), as it misplaces it in the sentence, which is carried over into the post-edited version. The second example shows a mistranslation of "enter" as an imperative, which could have been rectified by inserting "to" (All you need to do is *to* enter...).

The first example displayed in Table 6.7 involves colloquial language and cannot be translated directly into German. It is the computer rather than a person that is hacked, which is where the MT system fails. The post-edited version rectifies the grammatical errors "hatte, dass" → "wurde" but keeps the person as the subject,

	segment	adequacy	fluency	TER
colloquialism 1				
ST	She had been hacked.			
MT	Sie hatte, dass ein Hacker.	1	1	
gloss	She had, that a hacker.			
PE	Sie wurde gehackt.	3	3	125%
gloss	She was hacked.			
colloquialism 2				
ST	Just about every email with images in the html code are blocks and I have to manually OK each time to see them.			
MT	Nur über alle E-Mails mit Images in der HTML-Code angezeigt werden blockiert, und ich muss manuell "OK" jedes Mal, sie zu sehen.			
gloss	Just over all emails with images in the html code are shown as blocked, and I have to manually "OK" every time, see them.			
PE	Fast alle E-Mails mit Bildern im HTML-Code werden blockiert und ich muss manuell "OK" jedes Mal, sie zu sehen.	3.5	1.25	50%
gloss	Almost all emails with images in the html code are blocked and I have to manually "OK" every time, see them.			

Table 6.7: PE examples for colloquialism

rather than changing it to the computer (ideally: ihr Computer wurde gehackt, gloss: her computer was hacked) which is (presumably) what the domain specialists penalised in their evaluation. The dilemma faced by the post-editors then was to either retain the subject and produce an unidiomatic sentence or to change the subject and deviate from the original sentence. The second example includes two colloquial structures "just about" and "to manually OK", both of which the MT system mistranslates. The post-edited version corrects the first but opts to retain the less idiomatic and ungrammatical construction that is present in the MT output (ideally: ich muss jedes Mal manuell bestätigen um sie zu sehen, gloss: I have to manually confirm every time to see them).

The most problematic MT issues have been touched upon in the examples that have just been presented, which are mistranslations, incomplete translations (e.g.verbs),

	segment	adequacy	fluency	TER
mistranslation 1				
ST	Both PCs are running Windows 7 with NORTON 360.			
MT	Sowohl PCs mit Windows 7 mit Norton 360.	1.67	1.67	
gloss	Both[a] PCs with Windows 7 with Norton 360.			
PE	Sowohl PCs haben Windows 7 mit Norton 360.	3.67	3	11%
gloss	Both[b] PCs have Windows 7 with Norton 360.			
mistranslation 2				
ST	Recently, my ROM corrupted and bricked my phone.			
MT	Vor kurzem hat mein ROM beschädigt und bricked meine Telefon.	1.33	1.33	
gloss	Recently, has corrupted my ROM and bricked my phone.			
PE	Vor kurzem wurde mein ROM beschädigt und sperrte mein Telefon.	4	3	27%
gloss	Recently, my ROM was corrupted and locked my phone.			

TABLE 6.8: PE examples for mistranslations

---

[a]'Both' here is the first part of the construction 'both...and' ('sowohl...als auch').
[b]'Both' here is the first part of the construction 'both...and' ('sowohl...als auch').

missing information, unidiomatic translations and awkward sentence structures. These issues were sometimes retained in the post-edited versions, which may have resulted from the post-editing guidelines (Chapter 4, section 4.4.10) that prompted the post-editors to reuse as much of the machine translated segments as possible. In Table 6.8, the first example shows a mistranslation that has been carried over from the machine translated output, while the second example shows that the mistranslation occurred during the post-editing step, as "bricked" has been translated as "locked". This mistranslation may be due to insufficient knowledge of the source language, since "bricked" had remained untranslated (and is also a colloquialism).

6.3.2 Post-Editing Issues

The following section focusses on problematic edits that occurred during the post-editing phase. The Table 6.9 deals with segments, which have been reformulated by the post-editor, sometimes to produce a more idiomatic sentence sometimes for less apparent reasons. These reformulations have been summarised under the heading of paraphrasing, as they essentially aim to express the same meaning in different words with varying degrees of proximity to the original segment.

	segment	adequacy	fluency	TER
paraphrase 1				
ST	I won't have the time for a week to go through that.			
MT	Ich habe nicht die Zeit für eine Woche durch.	1.33	1.67	
gloss	I don't have the time for a week through.			
PE	Ich habe keine Woche Zeit für den Vorgang.	1.67	3.67	53%
gloss	I don't have a week's worth of time for this process.			
paraphrase 2				
ST	What is Norton called?			
MT	Was ist Norton heißen?	2	1.67	
gloss	Which is Norton call?			
PE	wie kann ich Norton in der Liste finden?	3.33	3	78%
gloss	How can I find Norton in the list?			

TABLE 6.9: PE examples for paraphrasing

In the first example of Table 6.9, the post-editor manages to fill in the blanks left by the MT system for 'go' and 'that' and comes up with a less literal but more idiomatic version. It is a close paraphrased approximation of the original sentence and yet adequacy is scored very low. The second example does not convey the meaning of the source segment. However, the context of the sentence renders the sentence correct, as the source segment is asking what Norton is called in the Task Manager list, which is spelled out in the post-edited version. This explains the higher score for adequacy. Example 2 also shows that the post-editor changed the sentence completely and retained

only "Norton".

	segment	adequacy	fluency	TER
short 1				
ST	Where should I go to get my activation key or do I need or hassle Google to get my apps list fixed			
MT	Wo sollte ich meinen Aktivierungsschlüssel oder benötige ich oder Aufwand Google meine apps Liste behoben?	1.33	1.33	
gloss	Where should I my activation key or do I need or effort Google my apps list fixed?			
PE	Wo kann ich mein Produkt aktivieren?	4	3	186%
gloss	Where can I activate my product?			
short 2				
ST	Yes, it's definitely the autorun feature.			
MT	Ja, es ist definitiv die selbstinstallierende Funktion.	4.75	2.75	
gloss	Yes, it is definitely the selfinstalling function.			
PE	Ja.	2.5	3.25	350%
gloss	Yes.			

TABLE 6.10: PE examples for short segments

Example 1 of Table 6.10 shows a drastically summarised sentence, where the score for adequacy was high but the score for fluency was not. It is arguable whether the sentence conveys the same meaning as the source and there are other examples (cf. Table 6.9) where the evaluators were not as lenient in their evaluations. Example 2, is drastically reduced, too. In context, it can be seen that the question was whether the "autorun feature" is the problem and this is the first sentence of the reply. The post-editor did not see it as necessary to repeat that information. This was penalised by the evaluators who were able to see the preceding and the following segment as context, but not the question conveying said information. Hence, not showing enough context in the evaluation may have influenced the evaluation of adequacy.

The post-editors produced unidiomatic or awkward segments, which could be due to lack of training as translators. The unidiomatic phrases presented in Table 6.11 are

	segment	adequacy	fluency	TER
unidiomatic 1				
ST	I am hesitant to use the Norton Power Eraser Tool.			
MT	Ich bin zögerlich zur Verwendung der Norton Power Eraser Tool.	4	3	
gloss	I am relucting to the use of the Norton Power Eraser Tool.			
PE	Ich bin zögerlich zur Verwendung des Norton Power Eraser Tool.	3	4	9%
gloss	I am relucting to the use of the Norton Power Eraser Tool.			
unidiomatic 2				
ST	Norton is simply making sure that nothing, not even Windows, can cause problems by disrupting Norton's operation.			
MT	Norton ist einfach dafür zu sorgen, dass nichts, nicht einmal Windows, Probleme verursachen können durch die Norton Vorgang.	2	1.33	
PE	Norton macht es einfach sicher, dass nichts, nicht einmal Windows, Probleme durch Unterbrechung von Norton-Operationen verursachen könnte.	3.33	2.33	52%
gloss	Norton is making it secure that nothing, not even Windows, could cause problems due to disrupting Norton's operations.			

TABLE 6.11: PE examples for unidiomatic segments

literal translations from English, the first of which was reused from the MT output and can be understood, whereas the second was produced by the post-editor and does not convey the meaning of the source segment.

Finally, carelessness of the post-editors was a source for low scores on occasion. In Table 6.12, examples displaying carelessness are shown. The first example shows two prepositions joined together "nachfür". The second example includes an extra letter "inoch". Both examples show that there are other translation issues that have not been corrected and show carelessness by the post-editor. The pattern of several careless mistakes in one segment was consistent throughout the post-edited segments.

However, as can be seen from the examples, it is often not only one problem but a

	segment	adequacy	fluency	TER
careless 1				
ST	Use that name when File Guard prompts for your admin password.			
MT	Verwenden Sie diesen Namen, wenn File Guard Anweisungen für Ihr Administratorkennwort.	2.67	2	
gloss	Use this name when File Guard prompts for your admin password.			
PE	Verwenden Sie diesen Namen, wenn File Guard nachfür Ihrem Administratorkennwort fragt.	3	2	31%
gloss	Use this name when File Guard prompts afterfor your admin password.			
careless 2				
ST	I would like to know if there is a fix for this, or maybe NORTON doesn't know about it yet.			
MT	Ich möchte wissen, ob es eine Lösung für dieses, oder vielleicht Norton ist es noch nicht kennen.	1.67	1.33	
gloss	I would like to know if there is a fix for this one, or maybe Norton is it not yet known.			
PE	Ich möchte wissen, ob es eine Lösung für gibt , oder es vielleicht bei Norton *inoch* nicht bekannt ist.	3.67	3	29%
gloss	I would like to know, if there is a fix for , or maybe it is not known *iyet* at Norton.			

TABLE 6.12: PE examples for carelessness

combination of source text, machine translation and post-editing issues that produces low quality output. The information presented above may be used to improve pre-editing, e.g. by enforcing explicitation. It can be further used for the purpose of training machine translation systems, e.g. expanding lexical coverage and including more training data with colloquialisms. Or it could be helpful in training post-editors, e.g. by reminding them to check their text again to avoid carelessness, to avoid literal translations from English, to pay attention to sentence structure especially and to reorder sentences, if necessary. In this experiment, the guidelines as presented in Chapter 4 section 4.4.10 may have contributed to the errors classified as being due to carelessness.

6.3.3 Limited Post-Editor Skills

In the following, the focus will be on post-editors for whom post-editing was expected to be more difficult, either due to a lack of or limited knowledge of the source or the target language. The post-editing results by the non-native speakers of German (two post-editors) were characterised by grammatical issues that range from minor (e.g. determiner, agreement) to severe (phrasal ordering), examples of which are displayed in Table 6.13. Another characteristic of their post-edited output was unidiomatic constructions. Some of the problems were based on the poorly formed source text. All these errors, however, had a minor impact on the adequacy score with an average of 4.2 for the non-native speakers, whereas fluency was rated as 2.4.

	segment	adequacy	fluency	TER
grammar 1				
ST	I want the other one back			
MT	Ich möchte eine andere haben.	3	3.75	
gloss	I want another one.			
PE	Ich möchte die andere wieder zu haben.	4.25	3	37%
gloss	I want the other one to have back.			
grammar 2				
ST	Or thirdly is there a way I can simply input the new product key and keep the existing installation with an extended time period?			
MT	Drittens gibt es eine Möglichkeit, kann ich nur die neuen Produktschlüssel und die vorhandene Installation mit einen längeren Zeitraum?	3.5	2	
gloss	Thirdly, is there a way I can only the new product key and the existing installation with an extended time period?			
PE	Oder drittens gibt es eine Möglichkeit, ich kann einfach nur den neuen Produktschlüssel eingeben und die vorhandene Installation mit einen längeren Zeitraum zu halten?	4.75	2.75	30%
gloss	Or thirdly, is there a way I can just type in the new product key and to hold the existing installation with an extended time period?			

TABLE 6.13: PE examples by non-native post-editors

The first example shows that non-native speakers of the target language (German) struggled with post-editing verbs correctly, especially in terms of verb tenses. The second example shows that they were inclined to transfer English sentence structures directly into German "ich kann...". The post-editor detected that the MT output structure was wrong but did not have enough knowledge of German to find the correct structure. Furthermore, this example displays other grammatical issues (preposition, verb form). This shows that the post-editor is able to understand the source and transfer the meaning correctly but introduces grammatical issues, which is then penalised by the domain specialists.

Monolingual post-editing (as performed by one participant) is mostly characterised by errors that are carried over from the machine translated version. This includes extra information and mistranslations, as shown in the example below, which the monolingual user had no possibility of verifying, due to the lack of knowledge of the source language, see Table 6.14.

	segment	adequacy	fluency	TER
extra information				
ST	I need to know asap please!! !			
MT	Ich benötige beides so schnell wie möglich bitte!!!	2.67	4	
gloss	I need both as fast as possible please!!!			
PE	Ich benötige bitte beides so schnell wie möglich!!!	3.67	2.33	9%
gloss	Please, I need both as fast as possible!!!			
mistranslation				
ST	However, NU should not have been deleting this			
MT	Aber NU sollten nicht gelöscht wurden	1	3	
gloss	But NU should not was deleted			
PE	Aber NU sollte nicht gelöscht werden.	2	1.67	29%
gloss	But NU should not be deleted.			

TABLE 6.14: PE examples by monolingual post-editors

Both examples show that the accuracy errors were carried over to the post-edited versions by the monolingual post-editor, who corrected language errors and implemented

changes that did not correct errors but did not introduce new errors either.

Fifty low-scoring segments remained unedited. Short segments that contained minor errors seemed to have been left unedited, with 26 segments containing 10 words or fewer. Longer segments $(15 - 33$ words) that remained unedited had an average of 5.1 errors per segment and often a poor sentence structure. The average amount of post-editing performed in the low-scoring category, as expressed by the TER scores, was 36% for the two non-native post-editors, 29% for the monolingual post-editor and 39% for the remaining 15 post-editors.

6.4 When Is Post-Editing Successful?

There were 47 distinct segments that scored 5 for both adequacy and fluency. Eleven (11%) of these scored 5 for adequacy and fluency before and after post-editing (TER = 0%); for these, post-editing thus consisted of reading and a decision that the segment did not require post-editing. The average score for the raw MT output was 2.5 for adequacy and 3.1 for fluency for the raw machine translated output in this group.

6.4.1 Abilities of the MT System

The examples presented in Table 6.15 show that the MT system was able to correct format errors (spaces) and to handle questions present in the ST.

	segment	adequacy	fluency
unchanged 1			
ST	Hi , thanks for the update.		
MT	Hallo, vielen Dank für die Aktualisierung.	5	5
PE	Hallo, vielen Dank für die Aktualisierung.	5	5
unchanged 2			
ST	What happens when a NAV license expires?		
MT	Was passiert, wenn eine NAV Lizenz abläuft?	5	5
PE	Was passiert, wenn eine NAV Lizenz abläuft?	5	5
unchanged 3			
ST	Thanks much.		
MT	Vielen Dank.	5	5
PE	Vielen Dank.	5	5

TABLE 6.15: Segments that did not require PE

For the high-scoring segments, there was a moderate correlation between adequacy scores of the raw machine translated output (0.68**) and the amount of post-editing performed as measured by TER (0-1 range) to achieve a segment scoring 5 and 5, and a strong correlation between fluency scores of the raw machine translated output (0.81**) and the amount of post-editing performed. Both correlations were highly statistically significant. This suggests that the evaluation of adequacy and especially fluency may be a good indicator for measuring post-editing distance and that AEMs, here TER, may be somewhat useful, albeit contradictory to previous observations.

Twenty segments (36%) fell into the category that required 0.1% to 35% of post-editing to achieve top scores for adequacy and fluency. Another 29 segments (52%) required between 36% and 92% of post-editing. Examples for both categories are presented below, with edits highlighted in red. It is evident that post-editing was successful whenever fixing a sentence required the changing of between one and three words, which included correcting a mistranslated word (light editing 1) or changing the number of an adjective and shifting a word (light editing 2), for example. It did not seem to matter what category the error belonged to, as long as fixing the segment required a low amount of editing. For examples with higher TER scores, it is evident that the post-editors rephrased more than they reused from the MT output (heavy editing 1-3). Thus, it may be worthwhile only showing MT output to potential community post-editors if the predicted post-editing effort is $\leq 35\%$. In Table 6.16, glosses are omitted for the MT output, as they do not capture the German errors that are specific to German (e.g. positioning of verb, number agreement). Instead, the edits are highlighted in red in the post-edited solution.

6.4.2 Successful Edits

	segment	TER
light editing 1		
ST	To work around the issue, you need to run a Quick Scan.	
MT	Um das Problem zu arbeiten, müssen Sie einen Quick Scan ausführen.	
gloss	To work the problem, you need to run a Quick Scan.	
PE	Um das Problem zu umgehen, müssen Sie einen Quick Scan ausführen.	8%
gloss	To work around the issue, you need to run a Quick Scan.	
light editing 2		
ST	You should now see the detailed tracking cookie information.	
MT	Sie sollten nun sehen, die detaillierte Tracking-Cookie Informationen.	
PE	Sie sollten nun die detaillierten Tracking-Cookie Informationen sehen.	33%
gloss	You should now see the detailed tracking cookie information.	
heavy editing 1		
ST	The only solution is for me to install the program and reinstall it again.	
MT	Die einzige Lösung ist für mich, installieren Sie das Programm und installieren Sie es erneut.	
PE	Die einzige Lösung für mich besteht darin, das Programm zu deinstallieren und erneut zu installieren.	53%
gloss	The only solution for me consists of deinstalling and reinstalling the program.	
heavy editing 2		
ST	Page freezes when loading, I can't scroll or interact with page at all.	
MT	Beim Laden der Seite Abstürze blättern, kann ich nicht auf all oder Interaktion mit Seite.	
PE	Seiten frieren beim Laden ein, ich kann überhaupt nicht scrollen oder mit der Seite interagieren.	65%
gloss	Pages freeze when loading, I can't scroll at all or interact with the page.	
heavy editing 3		
ST	Just to let you know when you purchase the upgrade to NMS from Android Market it will not show in your Norton Account product list.	
MT	Nur, damit Sie wissen, wann Sie das Upgrade auf NMS von Android Market wird es nicht in Ihrem Norton Account Produkt anzeigen.	
PE	Zu Ihrer Information, das NMS-Upgrade, wird nicht in Ihrer Norton Account-Produktliste erscheinen, wenn Sie das Upgrade über Android Market erworben haben.	92%
gloss	Just to let you know, the NMS upgrade will not show in your Norton Account product list, if you purchased the upgrade through the Android Market.	

TABLE 6.16: Segments that were post-edited successfully

6.5 Discussion of RQ2

This section will focus on answering RQ2[4] from a qualitative perspective.

In terms of segment statistics, it is evident that short segments influence fluency scores positively, supported by an average length of 14 words in low-scoring and 10 words in high-scoring post-edited segments. Furthermore, accuracy errors seem to influence adequacy and fluency equally, while language errors seem to influence fluency predominantly. With regard to segment characteristics identifying issues in the ST, the MT process and the post-editing process, Table 6.17 summarises the main problem characteristics that influence post-editing quality negatively.

problem	influencing
ST issues	
elliptical sentence structure	fluency
complex sentence structure	fluency
no explicitation	fluency
colloquialism	adequacy
mistranslation	adequacy
PE issues	
unidiomatic phrases	fluency, adequacy
carelessness	fluency, adequacy
paraphrasing	adequacy
shortening	adequacy
non-native	
language errors (predominantly grammar)	fluency
monolingual	
accuracy errors	adequacy

TABLE 6.17: Summary of unsuccessful post-editing

Successful post-edited sentences required mostly less than 35% post-editing (as measured by TER). Frequently, these sentences had more than one post-editor achieving top scores for both adequacy and fluency, while it was only ever one post-editor who

---

[4]What are segment characteristics and post-editing behaviour that render post-editing successful/unsuccessful?

achieved 5/5 for segments requiring considerably more post-editing. These segments were post-edited successfully, mostly by being rewritten, and required post-editors to be motivated enough to perform all edits required.

Complex or elliptical sentence structures and colloquialisms in the ST hampered correct MT output. For low-scoring segments, the ST problems could be addressed by expanding on the pre-editing rules, and by improving the MT output by adding colloquialisms and (technical) terms that were not yet included in the training data.

6.6 Concluding Remarks

This chapter approached quality of the post-edited content from a qualitative viewpoint. Low-scoring and high-scoring segments were identified and examined for segment characteristics both statistically and in regards to post-editing behaviour.

It was found for the high-scoring segments that rating fluency on a 5-point Likert scale was strongly correlated to the TER score, i.e. the post-editing distance. The principal focus was on the low-scoring segments, revealing that a lot of issues could be prevented before the machine translation step. Problems arising during post-editing could be rectified by targeting the specific issues outlined in the guidelines that are presented to the post-editors. Another approach to eliminating errors would be to introduce collaborative post-editing, i.e. several post-editors editing the same text, as explored by Tatsumi et al. (2012). The discussion presented in this chapter showed that quality of post-edited content is complex in an online community scenario and may not be fully portrayed by quantitative quality indicators only. While it is important to establish quantitative quality scores (cf. Chapter 5), it is vital to identify what ST

characteristics or edits caused them (see this chapter, Chapter 6) and to consider the

profile behind those editors, as will be presented in Chapter 7.

Chapter 7

Results III - Impact of Post-Editor Profile on Quality

7.1 Introduction

Subsequent to analysing the post-edited data on a quantitative and qualitative level, Chapter 7 aims at aligning the quality findings with the data collected on the post-editor profiles, to address RQ3:

RQ3) *Do elements of the post-editor profile serve as predictors for the quality of the post-edited output?*

In order to abstract from the individual post-editors and to take a step towards developing a theoretical basis for lay post-editing competence, some aspects measured for the post-editor profile are then used to test the Translation Competence model developed by Göpferich in order to address RQ4:

RQ4) *Is Göpferich's Translation Competence model applicable to community post-editing?*

The data collected on the post-editor profile will be used to identify any correlations between the profile and the quality of the post-edited output. The focus will be on the error annotation and the domain specialist evaluation. As the domain specialist and the community evaluation are strongly correlated, only the domain specialist evaluation is considered here, as it covers content (44%) for each post-editor equally.

Firstly, this chapter presents participant attitudes towards MT in the Norton Community and the motivation to post-edit and any correlations with the quality scores. Secondly, the focus will be on any potential correlations between the self-reported importance of grammar, content, spelling and punctuation in the Norton Community and the quality scores. Thirdly, the self-reported language competence for English and German will be explored in relation to the quality scores. These three variables will give some indication of whether self-reporting may be used here to help to predict PE quality. Fourthly, domain competence will be explored with reference to user statistics extracted from the Norton Communities and any correlations will be identified. Finally, the psychomotor competence will be investigated in relation to the quality scores. In summary, the independent variables that were identified as potentially influencing post-editing quality are: 1) importance of grammar, content, spelling and punctuation in the community, 2) attitude to MT and motivation to post-edit, 3) SL and TL skills, 4) domain knowledge and 5) psychomotor skills (cf. Chapter 4, section 4.2).

Figure 7.1 visualises these variables. While the "Attitudes" and the "Importance" (in light grey) are relevant to RQ3 only, TC1-TC3 (in dark grey) are aspects of the post-editor profile that are also used to answer RQ4 and correspond to sub-competences of
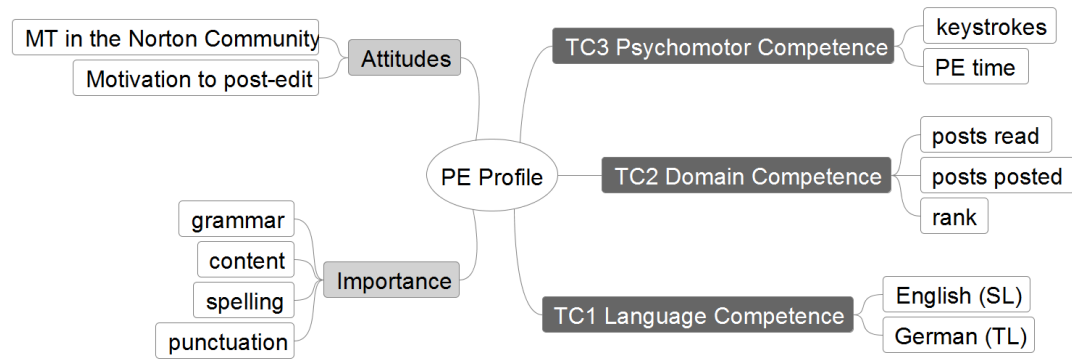
FIGURE 7.1: Overview of post-editor profile

Göpferich's Translation Competence model, as described in Chapter 2, section 2.2.1.2.

The data for each aspect of the post-editor characteristics will be presented, followed by any potential correlations and a discussion of the corresponding hypotheses. With this chapter, we hope to take a step towards identifying an ideal post-editor profile for community post-editing, i.e. a post-editor who corrects the majority of errors and performs well in the domain expert/community evaluation.

7.2 Attitude to MT and PE

The self-reported attitudes towards the usefulness of MT in the Norton Community and the self-reported motivation to post-edit machine translated posts were measured on a 5-point Likert scale before and after the post-editing experiment. An English translation of the German question is displayed below (see Appendix B for the German version).

Please indicate your opinion on the following statements on a scale of 1 to 5, where 1 = do not agree, 2 = rather not agree, 3 = neither nor, 4 = rather agree, 5 = agree.
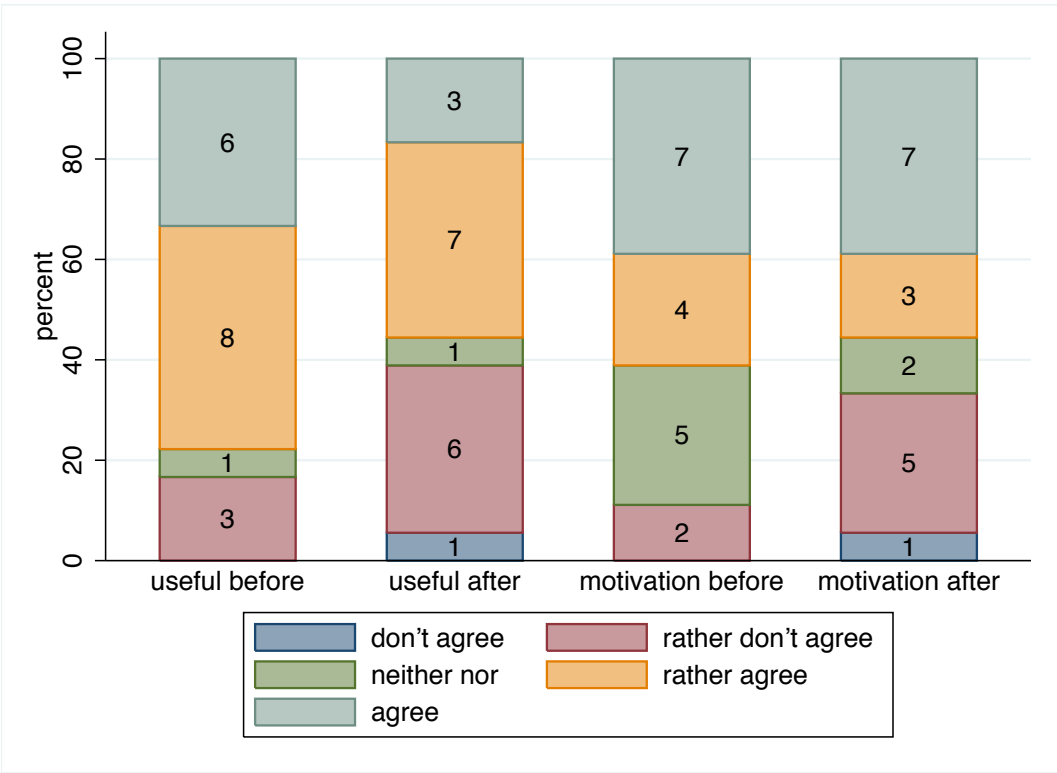
FIGURE 7.2: Attitudes: usefulness of MT/ motivation to post-edit content before and after the experiment

A) Machine Translation can be useful in the Norton Forum.

B) I am motivated to edit machine translated posts.

As can be seen in Figure 7.2 the attitude towards usefulness of Machine Translation in the Norton Community and the motivation to post-edit machine translated posts largely remained the same before and after the post-editing task. A slight decrease (not statistically significant) is perceivable in both after the post-editing task, however.

Table 7.1 presents the correlations between the usefulness of MT/the motivation to post-edit and the quality scores of the domain specialist evaluation (adequacy and fluency) and the error annotation (Total errors, Accuracy errors, Language errors and Format errors) with * as statistically significant and ** as highly statistically significant. Spearman's rho is used to compensate for the low number of participants and for

the data, which was not normally distributed. As can be seen, there are predominantly moderate correlations. While it may seem contradictory that there are negative correlations between these attitudes and the domain specialist evaluation and positive correlations for the error annotation, that is not the case. It is true for the domain evaluation that the higher the score is, the better the quality. For the error annotation, a higher score means that more errors remain in the post-edited output, i.e. the lower the quality is. It may be concluded from this table that the more the participants agreed that MT was useful and that they were motivated to post-edit, the lower their fluency scores and the more accuracy errors remained in the post-edited output. Generally, this means that the more positive they were during self-reporting, the lower the output quality was.

	usefulness	motivation
adequacy	-0.35	-0.38
fluency	-0.52*	-0.60**
total errors	0.47	0.54*
accuracy errors	0.52*	0.77**
language errors	0.38	0.33
format errors	0.28	0.05

TABLE 7.1: Spearman's rho between attitudes towards MT and motivation to post-edit and quality measures

With this information, the first hypothesis for RQ3 can be addressed:

H3.1$_0$) *There is no correlation between the reported attitudes towards machine translation and post-editing and the quality of the post-edited content.*

H3.1$_1$) *There is a correlation between the reported attitudes towards machine translation and post-editing and the quality of the post-edited content.*

While not all correlations were statistically significant, it emerges that there is indeed a correlation between these attitudes and the quality as measured by the error annotation and the domain specialist evaluation. $H3.1_0$ can be rejected as more positive self-reported attitudes result in poorer quality. This seems to be a paradoxical result, as one would expect that, the more positively post-editors see MT and the more motivated they are, the higher the quality of their post-edited content is. However, it could point towards a discrepancy between what post-editors say and do. It demonstrates that self-reporting needs to be used cautiously for measuring attitudes and may point towards an underestimation of post-editing effort.

7.3 Error Tolerances

In order to determine what type of errors would potentially have a negative impact on the perception of machine translated or post-edited content in the forum, the self-reported importance of content, grammar, spelling and punctuation was rated.[1] Figure 7.3 shows how participants rated the importance of grammar, content, spelling and punctuation in the Norton Communities before and after the experiment.

Content ('Inhalt'), defined here as the actual information delivered that the forum member is searching for, was rated as most important both before and after the experiment. The answers were concentrated on "rather important" and "important", which distinguishes it from the other categories. This result supports the assumption that content is crucial in the communities, if a user is to follow instructions to solve a problem. "Grammar" was rated as of second importance before and after the experiment

---

[1]importance as rated from 1-5, with 1 = not important, 2 = rather not important, 3 = neither nor, 4 = rather important and 5 = important
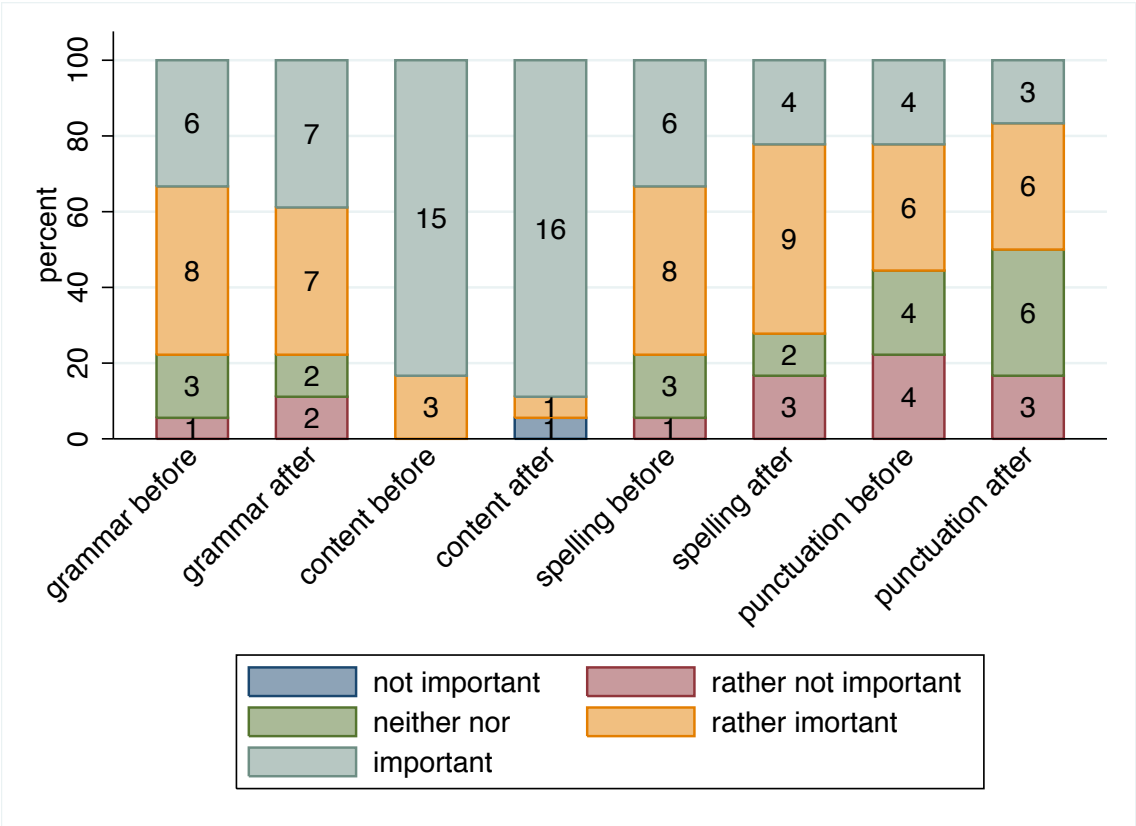
FIGURE 7.3: Self-reported error tolerances

on average followed by "spelling". Punctuation was rated as least important in the Norton Community both before and after the experiment, which is interesting given that a lot of post-editing focusses on fixing punctuation (see e.g. O'Brien and Roturier 2007). Figure 7.3 also demonstrates that the post-editing experiment had little to no impact on the error tolerances represented by the importance of content, grammar, spelling and punctuation in the Norton forum.

To test whether there is a correlation between self-reported error tolerances and PE quality, Spearman's rho was calculated for the importance of grammar, content, spelling and punctuation and the quality measures (error annotation and domain specialist evaluation), as presented in Table 7.2.

Weak correlations are predominantly present between the importance of the aspects

	grammar	content	spelling	punctuation
adequacy	0.03	-0.06	-0.15	-0.34
fluency	-0.13	-0.22	-0.04	-0.23
total errors	0.01	0.35	-0.04	-0.23
accuracy errors	0.26	0.16	0.18	-0.01
language errors	-0.27	0.42	-0.27	-0.20
format errors	0.05	-0.12	0.04	-0.03

TABLE 7.2: Spearman's rho for importance of grammar, content, spelling and punctuation and PE quality

and the post-editing quality, none of which are of statistical significance. On the basis of this information, the following H3.2$_0$ could not be rejected:

H3.2$_0$) *There is no correlation between the reported importance of grammar, content, spelling and punctuation and the quality of the post-edited output.*

H3.2$_1$) *There is a correlation between the reported importance of grammar, content, spelling and punctuation and the quality of the post-edited output.*

There is no correlation between the reported importance of grammar, content, spelling and punctuation and the quality of the post-edited output. This is the case for the domain specialist evaluation (adequacy and fluency) and for the error annotation results both on a general and a more detailed level. This was confirmed for the types of errors corrected and the attitude towards errors (cf. Appendix M for an empirical breakdown). It is evident that self-reporting did not prove to be useful here.

7.4   Language Competence

Figure 7.4 shows the self-reported English competence (reading) and the self-reported German competence (writing) according to the Cedefop description (2011).

While English language competence ranged from categories A-C with a tendency towards C1 and C2, the two most advanced categories, the German language competence was concentrated on categories B and C, with 22% of the participants rating themselves as B, one of whom was a non-native speaker of German. The second non-native speaker rated himself as C1.
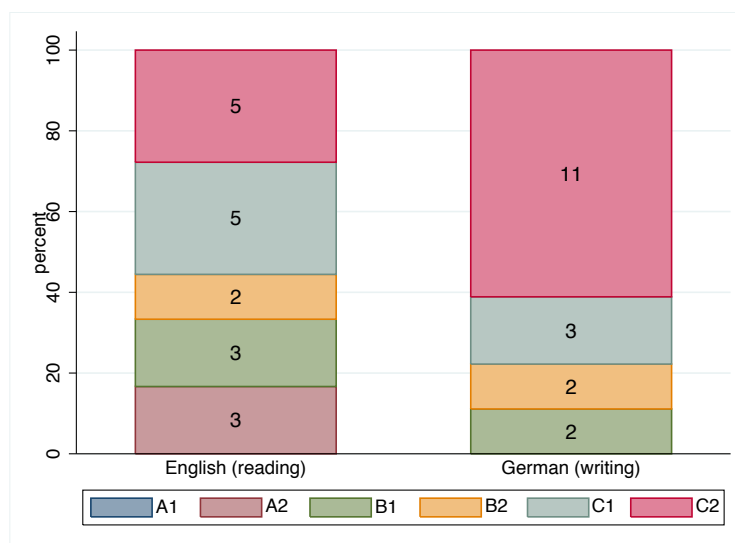
FIGURE 7.4: Reported language competence for English (reading) and German (writing)

Table 7.3 displays the correlations between the participants' self-reported language competences and the PE quality with * as statistically significant. Again, the apparent discrepancy between positive and negative correlations is due to a higher adequacy/-fluency score indicating higher quality and a higher number of Total/Accuracy/Language/Format errors indicating lower quality output. While correlations range between negligible to moderate, only one of them is statistically significant. The number of Accuracy errors is moderately negatively correlated to the self-reported knowledge of English, i.e. the higher the knowledge of English (indicated), the fewer Accuracy errors remain.

In the light of these observations, the following hypotheses may be discussed:

	knowledge (de)	knowledge (en)
adequacy	0.00	0.39
fluency	0.37	0.43
total errors	-0.27	-0.37
accuracy errors	-0.33	-0.56*
language errors	-0.27	-0.16
format errors	-0.17	-0.29

TABLE 7.3: Spearman's rho between language competence and quality measures

H3.3$_0$) *There is no correlation between the reported communicative competence in English and German and the quality of the post-edited content.*

H3.3$_1$) *There is a correlation between the reported communicative competence in English and German and the quality of the post-edited content.*

As there is only one statistically significant correlation, H3.3$_0$ could not be rejected with confidence. However, it points towards the fact that SL skills are directly linked to the number of Accuracy errors and possibly adequacy. It indicates that the measure chosen, self-reporting of language competences, is not sufficient here. We hypothesise that with a more apt method, such as a language test, correlations may appear more clearly.

7.5 Domain Competence

7.5.1 Ranking

Generally, there are two types of online community members, "posters", users sharing their experience within the online community and "lurkers", consuming this content without providing any content (Schlosser 2005). Studies have found that it is most

often a small percentage of users who provide the largest share of content (e.g. Ling et al. 2005). An online community is as much a community as offline communities. It is important to keep in mind that users' behaviour is tied to the individual, i.e. it is based on the users' desires, cognitions and abilities to "interpret and interact with their environment" (Bishop 2007, p. 85) and is therefore largely unpredictable.

McAuley, O'Conner and Lewis (2012) developed and tested the community analysis tool, Petri, while drawing on knowledge gained by studying the German Norton Community in detail. Petri operates using a clustering technique based on eight primary features, such as posts and kudos received, and seven supplementary features, such as number of logins and kudos given. Apart from the large number of passive users, lurkers, they identified four different types of (active) users, i.e. posters:

1. Pragmatist [...]: Users that ask a question and then receive an answer. Generally, they do not contribute to other threads that they have not initiated. They receive very little kudos and are mainly in the community to seek an answer to a question. These users make up the majority of active users in the community.

2. Peer [...]: Users that contribute to threads that they have not initiated. However, they receive little recognition from their peers, by way of kudos, and have lower contribution rates than the apprentice culture. This group shows some potential.

3. Apprentice [...]: Users that make a recognisable contribution to the community, participating in more threads than they initiate, thereby contributing more answers than questions. They have also begun to receive recognition in the form of kudos, views and accepted solutions from the

rest of the community.

4. Guru [...]: Finally, gurus contribute the most to the community. They are highly knowledgeable and skilled users that receive the most peer recognition in the form of the kudos, views and accepted solutions. (McAuley, O'Conner and Lewis 2012, 18f)

While this categorisation provides in-depth insight into the community, it could not be used for the purpose of this thesis, as the tool was operated with a model of the community, one year prior to the start of the post-editing study, which renders the community member information outdated. However, it informed the decision to abandon the somewhat inconsistently assigned ranks in the Norton Community and to combine them as follows. The ranks that are captured here are SYMC, guru and other. SYMC stands for Symantec employees, who are monitoring the forum and are active in the forum whenever necessary.

7.5.2 Community Member Activity

The information presented in Table 7.4 was extracted from the German and English Norton Communities in order to establish the post-editors' domain knowledge and to revisit their language competences. Domain knowledge is measured in replies posted and messages read in the German and the English forum, as well as the post-editors' rank in the forum.

Table 7.4 shows that apart from A2, all post-editors were active members of the Norton Communities. It clearly shows the difference between gurus and the groups of Symantec employees and other users of the Norton Community, i.e. the gurus are far more

	replies (de)	messages read (de)	replies (en)	messages read (en)	rank
A1	97	3223			other
A2					
A3	119	3841			other
A4	38	1452			other
A5	21	3631			SYMC
A6	42	1498			SYMC
A7	293	6128			other
A8	530	16227	5514	414921	guru
A9	24	1518			other
B1	952	31105	6452	182068	guru
B2	19	2748			other
B3	161	8356		909	SYMC
B4	74	2009			SYMC
B5	882	23289	13	2443	guru
B6	16	679			other
B7	67	2348			other
B8	265	15769	52	1612	guru
B9	3	196			other
avg guru	657	21598	3008	150261	
avg SYMC	75	3874	0	227	
avg other	68	2213	0	0	

TABLE 7.4: Domain competence as extracted from the Norton Community

engaged on average both in the German and the English-speaking Community. All four gurus (A8, B1, B5 and B8) are members of the English Norton forum, with the first two being more active on the English forum than the German forum. The engagement for Symantec employees and members of the community holding other ranks are comparable.

To return briefly to language competence, we will investigate whether this information is related to the self-reported language competence skills. Participants A8 and B1 posted more in the English Norton Community than in the German Norton Community. Both indicated the same value for their English and German competence, C2 and C1 respectively. While A8 is a bilingual speaker of both English and German, B1 is a non-native speaker of German and English. Investigating the relation between the

self-reported level of English and the level of activity in the English Norton Community further revealed that seven (A4, A5, A6, B2, B3, B4, B6) of the nine participants who rated their English skills as C1 or C2 were not active in the English Norton Community. The other two participants (B8, B5) who were active in the English Norton Community indicated their English skills as B1 and B2. This is supported by the fact that there are no statistically significant correlations between the reported language skills and the activity in the respective Norton Communities. The self-reported language competences are not drivers of the user activity in the English forum.

Table 7.5 displays correlations between the domain knowledge and the PE quality. It is remarkable that the correlations are predominantly negligible or weak and that none of the correlations are statistically significant.

	posted (de)	read (de)	posted (en)	read (en)	rank
adequacy	-0.11	-0.05	-0.07	-0.03	0.06
fluency	-0.21	-0.24	-0.31	-0.32	-0.07
total errors	0.00	0.10	0.12	0.13	-0.07
accuracy errors	0.23	0.17	0.17	0.16	-0.05
language errors	-0.04	0.24	0.08	0.13	0.01
format errors	-0.25	-0.14	-0.02	-0.02	-0.17

TABLE 7.5: Spearman's rho between domain knowledge and quality measures

H3.4$_0$) *There is no correlation between the domain competence and the quality of the post-edited content.*

H3.4$_1$) *There is a correlation between the domain competence and the quality of the post-edited content.*

Referring to Table 7.5, Hypothesis H3.4 may be addressed as follows. H3.4$_0$ could not be rejected, due to a lack of meaningful and statistically significant correlations. Domain competence, as measured in this experiment, and the quality of the post-edited content are not correlated for community post-editing for this experiment. While the

methods of eliciting this data was objective, it points to the fact that there are other (possibly psychological) characteristics, such as motivation, which were not controlled for or measured in this experiment, which overrule Domain Competence.

7.6 Psychomotor Competence

In order to measure the psychomotor competence as defined by Göpferich, the average time (in seconds) and the average number of keystrokes per task were recorded for each post-editor, as displayed in Figure 7.5.

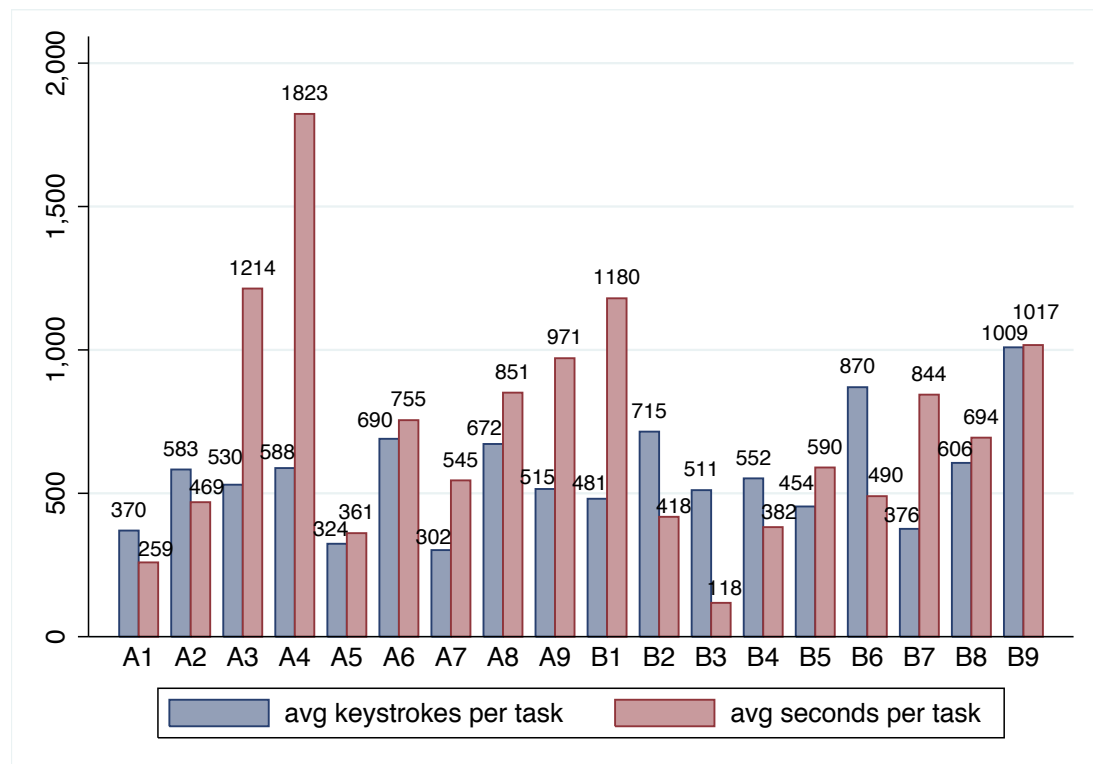FIGURE 7.5: Psychomotor competence in keystrokes and seconds per task

It is evident that both the number of keystrokes and the seconds spent working on one task at a time varied greatly across the post-editors. While the minimum number of keystrokes was 302 per task and the minimum time spent was 118 seconds (approx. two minutes), the highest number of keystrokes was 1,009 and the maximum time spent was

1,823 seconds (approx. 30 minutes). The overall average number of keystrokes was 564 and the overall number of seconds spent per task was 415 (approx. 7 minutes). It has to be considered here that the average length of a task was 175/170 words (English ST/German MT output). The average number of words edited per hour would be 1,518 (with 346 as the minimum and 5,339 as the maximum number of words edited per hour). There are only weak correlations between the average number of keystrokes and seconds per task and the scores for adequacy or fluency, as shown in Table 7.6, none of which are statistically significant. The same can be observed for the results of the error annotation.

	avg keystrokes	avg seconds
adequacy	0.18	-0.28
fluency	0.27	-0.11
total errors	-0.03	-0.19
accuracy errors	-0.12	0.08
language errors	-0.27	-0.27
format errors	-0.13	-0.11

TABLE 7.6: Correlations between quality and psychomotor competence

	avg keystrokes	avg seconds
adequacy	0.06	-0.55*
fluency	0.04	-0.10
total errors	0.29	-0.29
accuracy errors	0.13	0.10
language errors	0.00	-0.37
format errors	0.05	-0.05

TABLE 7.7: Correlations between quality and psychomotor competence, controlled for language skills

Table 7.7 shows correlations for the psychomotor competence as measured by keystrokes and number of seconds spent excluding the monolingual post-editor and the non-native speakers of the TL with * as statistically significant. These were excluded in the sample in order to control for the language skills, i.e. to include only native

speakers of German with some level of English to ultimately render the experimental design more rigorous. As can be seen, the trend that there is no correlation between keystrokes and adequacy/fluency as measured and between post-editing time and fluency was confirmed. However, there appears to be a moderate negative correlation between the post-editing time and the adequacy scores, i.e. the longer spent on the task, the lower the adequacy scores. For the other error annotation, the correlations change only marginally.

The data presented may be used to address H3.7:

H3.7$_0$) *There is no correlation between the psychomotor competence and the quality of the post-edited content.*

H3.7$_1$) *There is a correlation between the psychomotor competence and the quality of the post-edited content.*

Generally, H3.7$_0$ could not be rejected, due to a lack of meaningful and statistically significant correlations. However, there seems to be a correlation that the longer a post-editor spends on a task, the lower the adequacy scores are. A possible explanation could be that the poorer their knowledge of the source language, the longer they spend reading and thinking about the meaning of a segment, without achieving the same results that a post-editor with more advanced language skills of the SL will achieve. There is no statistically significant correlation between the self-reported knowledge of English and the time spent post-editing. As it has been established that the self-reported knowledge does not correlate with the PE quality, PE time may possibly be an indirect indicator of SL skills. It can be further concluded that the number of keystrokes are not a good indicator of quality, in contrast to the TER values, as presented in Chapter 6, section 6.4.2.

7.7 Discussion of RQ3

Based on the data collected, the characteristics of the post-editor profile measured individually using methods of self-reporting and forum statistics do not serve as predictors for the quality of the post-edited data. On a granular level, however, positive self-reported attitudes towards the usefulness of MT and the motivation to post-edit seem to be predictors for lower PE quality and long post-editing time, also seem to be predictors for lower PE quality. These observations show that there may be characteristics of the post-editor profile that serve as predictors for quality levels but that it is difficult to measure them reliably. Some important first steps have been taken towards identifying what these are and how they may or may not be measured. These characteristics are possibly inter-connected and may develop in a staggered manner and in a complex relationship, as identified by Göpferich in the light of Dynamic Systems Theory.

7.8 Discussion of RQ4

The data and observations presented in this chapter may be used further to partially explore the applicability of Göpferich's Translation Competence model for community post-editing. It needs to be noted that only three of the six competences outlined by Göpferich were tested in this experiment. This experiment differs from Göpferich's experiment in that hers was a longitudinal study of the development of translation competence in translation students, while this experiment was a short study measuring current competences in community post-editors without training in translation and post-editing selectively. The competences that were measured here are:

1. *Communicative competence in at least two languages* (here SL and TL) - self reported in pre-task survey

2. *Domain competence* - user statistics from the Norton Communities

3. *Psychomotor competence* - quantitative data collected during PE process

In regards to the communicative competence in English and German, it can be concluded from the data presented in section 7.4 that language skills are not linked to the post-edited quality. In regards to post-editing competence, good knowledge of the SL is required to eliminate accuracy errors during the post-editing process. With the data from this study, a relationship between TL skills and the fluency of or Language errors remaining in the post-edited output could not be confirmed. However, the sample was small and it could not be controlled for all variables of the post-editor profile.

Domain competence, as measured by the user statistics presented, is not a dominant or determining factor for high post-editing quality, as there were no correlations between the domain competence and the post-edited quality (see section 7.5). It may be the case that it is a quality that becomes significant once other competences (e.g. language skills) have been developed. This notion is based on the DST, as described also by Göpferich. In order to test this, data for more post-editors over a long period of time would be required.

Initially, psychomotor competence, as measured in keystrokes and seconds spent per task, did not seem to be correlated with post-editing quality (see section 7.6). After a certain degree of controlling for the language skills by only including post-editors with German as their native language who post-edited bilingually, i.e. with reference to the ST, it emerged that higher post-editing times serve as a predictor for lower adequacy scores. We hypothesise that rather than the time, it is the source language skill that is

visible in this correlation, which moves away from psychomotor competence. A better developed psychomotor competence here did not seem to be a predictor for quality. Again, it may be the case that this is a secondary competence, the effect of which can be measured when primary competences are fully developed and/or controlled for.

For community post-editing, Göpferich's Translation Competence model does not appear to be a helpful starting point as a theoretical basis, as some competences may never be developed (e.g. tools and research competence), different post-editors may possess a different skill set or possess the same skills with different strengths. They may be able to compensate for those skills that they do not possess and may be able to produce output of the same quality as others with different skills. We anticipate a competence model for lay post-editing to be more flexible than a model for post-editing for a professional translation environment. This is also based on additional competences or factors influencing the post-editing competence. These could be of a psychological nature, such as motivation or interest in the community. As stated in Chapter 2, section 2.2.1.2 Göpferich points out the importance of psychological factors that she does not study explicitly in her experiment, but which she includes in her model, such as the "psycho-physical disposition" of the translator and the "working conditions (e.g. time pressure)" (Göpferich 2009b, p. 21). In community post-editing, these may gain more importance than in a professional scenario. These present the challenge of finding appropriate methods to measure them. Future studies with more participants would be able to shed light on the impact of the individual competences described. A study with more emphasis on these extra factors that determine the "employment of the sub-competences [...] and their central control" (Göpferich 2009b, p. 21) would be helpful in determining their exact role.

7.9   Concluding Remarks

This chapter presented the data sets collected on the post-editor profiles and aimed at matching these with the post-edited quality obtained. We found that the post-editor profiles were very diverse. Despite this diversity, we found that self-reporting attitudes towards MT and PE were negatively correlated with post-editing quality. Furthermore, longer PE times served as predictors for lower adequacy scores, which we hypothesise to be an indicator of low (or insufficient) SL skills. In general, the data obtained in this study did not yield statistically significant correlations allowing for a clear connection between the post-editor profile and PE quality, represented by statistically significant correlations. In regards to a post-editing competence model, Göpferich's model was not found to be useful as it stands for the competences examined in a community post-editing scenario, where different aspects (e.g. of a psychological nature) may be more prominent than in traditional translation scenarios.

Chapter 8

Conclusion

8.1 Research Aims

The research presented here was motivated by the need to investigate how feasible post-editing by lay post-editors is in order to translate useful user-generated content effectively in online communities.

Post-editing has been researched extensively from a bilingual perspective and very little from a monolingual perspective. The literature would indicate that there are still gaps to fill in monolingual PE studies. The literature also reveals that monolingual post-editing studies are difficult to design and conduct for reasons of recruitment and motivation. Previous research by the present researcher has revealed that monolingual post-editing studies are extremely difficult to conduct in the Norton Communities due to a shortage of purely monolingual people, i.e. those people who do not have any knowledge of the source language. This is coupled with participant irritation that stems from withholding the source text. Abstracting from the monolingual/bilingual

discussion, lay post-editing, a field that is only just emerging, was considered a better fit for the context at hand. Little research has been conducted with lay post-editors, for example domain specialists. No post-editing studies have been conducted in already existing online communities of lay post-editors.

The first central concern was post-editing quality within the context of the online Norton Community, both on a quantitative and a qualitative level. Measuring PE quality involved two traditional top-down approaches, an error annotation based on de Almeida (2013) and a domain specialist evaluation of adequacy and fluency (LDC 2005). In addition, a novel perspective to quality was explored by devising a bottom-up approach, i.e. eliciting feedback on the final translation product from its end-users. A subset of the machine translated and post-edited data was judged by the Norton Community members on fluency on a 5-point Likert scale.

The second central concern was rooted in the concept of *lay* post-editing, namely the post-editor profile and its relation to PE quality. The post-editor profile involved the participants' self-reported attitudes towards machine translation in their community, their self-reported motivation to post-edit and the rated importance of grammar, content, spelling and punctuation within their community. It further involved their self-reported language competences (SL and TL), their domain competence as measured by user statistics and their psychomotor skills as recorded during the post-editing process.

The third central concern was taking a step towards adapting or developing a lay post-editing competence model. Due to a lack of such a model, an existing translation competence model (Göpferich 2013) was chosen to establish whether it may be used as a basis for a lay post-editing competence model. With the data from the full-scale

experiment and ultimately a post-editing competence model, it was hoped to identify helpful post-editor characteristics to optimise PE quality.

8.2 Findings

8.2.1 PE Quality QUAN - RQ1

The data collected on the post-editing quality revealed that all lay post-editors improved the raw MT output with statistical significance in the error annotation and for all but one participant in the domain specialist evaluation.

As was to be expected, the quality scores varied greatly across the lay post-editors. The error correction rate ranged from 30% to 87%. In the domain specialist evaluation, between 39% and 87% of all post-edited segments achieved the highest score ("5 - all meaning") for adequacy compared to an average adequacy of 35% for raw MT output. For fluency, between 18% and 54% of all segments reached the top score ("5 - perfect") compared to an average fluency of 11.5% for raw MT output. While there is still room for improvement in post-editing quality, it is important to consider that previous translation and post-editing studies have demonstrated that professional translations do not tend to achieve a score of 100% (e.g. PE quality: 82.67% (de Almeida 2013, p. 186); reference translations: 63.5% (Koehn 2010, p. 544); translation quality: 90% and PE quality: 95% (Plitt and Masselot 2010, p. 15); translation quality: 73% (Sinaiko and Brislin 1973, 332f)).[1] Therefore, the post-editing quality scores obtained here are an encouraging result for community post-editing.

---

[1]All percentages were derived from the data presented on the pages indicated.

8.2.2 Evaluation Approaches Compared - RQ1

The three evaluation approaches measured the concept of quality from three different perspectives (number and type of errors, fluency and adequacy, fluency only) and were thus not necessarily expected to correlate. Moderate correlations indicated that the more errors are present, the lower the evaluation scores for both the domain specialist and the community evaluation. This finding suggests that evaluators do not necessarily have to come from a pool of trained translators/linguists and that online communities, the end users of the content, may be well-placed to evaluate translation quality for their own purposes. This is a somewhat unexpected positive outcome. Considering the bias, variance and unreliability that trained linguists introduce in the evaluation of MT and PE content (cf. Denkowski and Lavie 2010, Lommel, Popovic and Burchardt 2014), we assumed this to be true for any evaluator and had expected the community evaluation to deviate greatly from that of the domain specialist evaluation. The variance, however, appears to have been levelled out by the numbers of evaluators. The existing body of research indicates that online (sub-) communities especially designed for translation are very well able to translate (and indeed evaluate) translated and post-edited content reliably. This has been shown in the examples of Facebook (Garcia 2010) employing a crowd-voting mechanism or the localisation of the browser Firefox (Drugan 2013, 166f) employing proof-reading by the crowd. With this thesis, an investigation of the ability to post-edit and evaluate is extended to online communities with primary purposes other than translation, without the creation of an explicit sub-community for translation. Concerns expressed regarding the evaluation process in such instances, e.g. discerning the best solution in clashing evaluations or malicious behaviour of a community evaluator (Drugan 2013, p. 169), were not found to be of relevance here.

8.2.3 PE Quality QUAL - RQ2

Quality was further examined on a qualitative level, in order to identify successful/unsuccessful post-editing behaviour. It revealed that the reasons for unsuccessful post-editing arise at different stages in the translation process. They were either carried over from the ST, occurred in the MT step or were introduced in the post-editing step. Accuracy errors were sometimes newly introduced but in the main were carried over from erroneous MT output. Language errors were often introduced by non-native speakers or remained uncorrected if they were "too obvious" for instance. For examples which required more editing than 35% (as measured by TER), no post-editing patterns could be detected. In general, short segments and segments requiring less than 35% of editing were most often corrected.

This suggests that content that is intended for machine translation may benefit from quality estimation or post-editing prediction techniques, in that only content that requires post-editing below a certain amount (such as 35%) would be displayed. If the translation could be understood without the missing text, the text requiring heavier PE could be discarded. Or, if a sub-community existed/ were to be built around translation/post-editing, the remaining content could be sent to a user with higher SL skills (or a professional translator), as has been suggested by Schwartz (2014).

8.2.4 Post-editor Profile - RQ3

The wide spread of quality scores was suspected to be rooted either in varying interpretations of the post-editing and evaluation guidelines, or in the post-editor profile. This leads to the discussion of the post-editor profile in relation to PE quality. We

found that in most cases, the post-editor profile characteristics were not correlated to the PE quality. This applied to the domain knowledge measured by user statistics and to self-reported language skills. A lack of correlation between domain knowledge measured by user statistics and the rank in the community (objective measures) was surprising. This suggests that domain knowledge may not be a dominant factor for PE quality or that other skills may have to be developed first in order for domain competence to fully unfold its effect on quality.

It was unexpected that language skills did not seem to be correlated to PE quality. It could be observed, however, that non-native speakers of the target language and monolingual post-editors performed at the lower end of the spectrum in the error annotation and the domain specialist evaluation step. Self-reporting is always subjective and may not have been an appropriate measure here, as it may render the identification of correlations pointless, if it is not consistent. However, self-reporting in rating the usefulness of MT in their community and the motivation to post-edit was correlated with statistical significance to the PE quality, in that "more positive" attitudes actually correlated with poorer PE quality.

Further, the time spent on a single task was negatively correlated with the PE quality, which is also a surprising result. This suggests that there is a certain time threshold after which PE productivity decreases as the post-editor is trying to solve translation problems which s/he cannot solve due to limited language skills for example. It has been emphasised that PE quality varies according to the post-editor/translator and often does not relate to profile characteristics/preferences measured (e.g. Teixeira 2014, Vieira 2014, Gaspari et al. 2014, de Almeida 2013, Koponen et al. 2012, Koehn 2010). The data gathered in this thesis suggests that PE quality may be rooted largely

in aspects of the post-editor profile that have not been measured here, as has also been previously suggested (e.g. Göpferich 2009a). Aspects of the post-editor profile that could be usefully measured next would be motivation to post-edit, the translator's self concept (Göpferich 2008) and a (search) tool competence adapted from Göpferich (2008), which for lay post-editing, I would anticipate to include online search engines and online (machine) translation services, for example.

Based on these observations, apart from giving preference to users with German as a native language and users with some level of English, we would not recommend post-editing for a group with particular characteristics only, such as gurus (super users), at this stage.

8.2.5 A Post-Editing Competence Model - RQ4

From the discussion of the user-profile, it follows that the Göpferich Translation Competence model is not useful in the context of lay post-editing as it stands. While only three of the sub-competences were tested in this thesis, it is unlikely that the other competences, which are competences that are traditionally acquired in official translation training, would yield insight as dominant competences into lay post-editing. We hypothesise further that additional competences or characteristics that are of peripheral nature in combination with others may be more dominant characteristics for lay post-editing, such as motivation coupled with language skills.

In short, community post-editing is a feasible concept, producing variable output, which is not dependent on the post-editor profile. Post-editing is more likely to be successful when the segments are short and the post-editing effort is $<= 35\%$ as measured by TER. Post-editing seems to require a different model, or a model with emphasis

on different sub-competences, than the TC model proposed by Göpferich that was developed for professional translation.

8.3 Contributions to the Field

As previously emphasised, this study is unique in that it approaches post-editing from a new perspective, that of volunteer post-editors of an online community with technology support as a primary purpose. This thesis not only deals with lay post-editors who are likely not to be trained in translation or may not have any knowledge of the source language, it further deals with a real life use case - post-editing in the Norton Communities. It takes up the idea of crowd-sourcing translations and post-editing solutions and tests it with a community that is already established. It bridges the gap between academia and industry, investigating post-editing from a translation perspective within the real world.

8.3.1 Online Communities as a Platform for Research

Apart from the obvious contribution that community post-editing is a feasible concept here, insights into the workings of and working with OCs are valuable contributions. Initially, the self-governed and very independent community with their own expectations and views was studied. Based on this, a post-editing interface and tasks that are useful for community members, rather than translators, had to be developed. A strategy of trial and error and several rounds of testing followed. The knowledge gained

from this experience is invaluable, as it is new and can be applied (modified if necessary) to other communities for future post-editing experiments in a field that is on the rise.[2]

The development of the methodology for this thesis exposed the challenges of working with an online community and with volunteers, especially in terms of recruitment and expectations that one should/should not have. Especially for a community with a primary purpose other than translation, one cannot expect members to join on their own account. They (regular users or super users) have to be led to the topic but need to explore and tease out on their own the implications and potential for themselves and the community. An active peer-to-peer relationship with the community is essential to successful recruitment for post-editors. To engage even one-time users or lurkers, anonymity, gamification, simplicity and concision seem to be key, as elicited from the community evaluation.

8.3.2 New Perspectives on Evaluation

Another major contribution lies in the evaluation design of the post-edited content, the scores and their implications. The topic of crowd-sourcing post-edited content and the evaluation thereof, i.e. quality control, is gaining importance especially in industrial circles. The journal article by Mitchell, O'Brien and Roturier (2014) that is based on this thesis and which focusses on the evaluation of post-edited content and the crowd-sourcing of evaluations by the end-users of the post-edited content, sparked a round-table discussion at the TAUS meeting in 2014. The evaluation process is versatile in its three components. The community evaluation is unique and innovative in the

---

[2]This experience has been shared by the present researcher in an industrial context at the TAUS Quality Evaluation Summit, hosted by Localization World on 4 June 2014.

field of Translation Studies. It breaks with the tradition of using translators/linguists as evaluators only and stresses the importance of giving a voice to the end-users of the translations.

As a consequence of working with community post-editors and evaluators, this thesis challenges long-standing assumptions about post-editing, such as: you need to be a translator or linguist to post-edit machine translated content. Or: you need to be a translator or linguist to evaluate post-edited content. It demonstrates that community members are willing and able to evaluate content reliably with regard to fluency.

In regards to comparing the three evaluation approaches, the following may be concluded. While different types of evaluation may yield correlating results on a macro-level, they most likely will yield differing results on a micro-level, as has also been observed by Sinaiko and Brislin (1973). This suggests that quality as a concept is not only complex but that the totality of it cannot be measured by choosing one approach. This is not surprising if you consider that different "groups" of subjects performing evaluations bring a different set of expectations to the texts.

Even individual characteristics within those groups affected great variance in the case of the domain specialist evaluation in the experiment presented here. This was particularly true between group A (3 evaluators) and B (4 evaluators) and their evaluation approach to adequacy, as well as variance within the two groups. The results for the community evaluation were much more stable in that regard but had at least eight evaluators per segment. While we are not in a position to pinpoint an exact threshold for a useful number of evaluators, it suggests that the previously assumed number of 3-4 evaluators that is commonly used in the field to evaluate translations is not adequate,

at least not for domain specialists. Thus, opening the evaluation to the community in question may not only be "easier", it may also yield more reliable results.

Going forward, this stresses the importance of selecting an evaluation approach that is appropriate for the context it is produced in or for and the type of quality that is required. Depending on these requirements, this may be done by choosing a single approach if a particular aspect of quality is to be examined. It may also involve combining several approaches, such as was done here or may be operationalised with the help of the DQF tool (TAUS 2013a) if a more comprehensive understanding of quality is required.

8.4 Methodology Revisited

The involvement with the ACCEPT project opened up opportunities to this thesis that would have been impossible otherwise. This included a holistic view of the ACCEPT project and all its components, including pre-editing, machine translation, post-editing, evaluation and the ability to leverage expertise and technology from project partners in all of these areas. This enabled the present researcher to design a series of experiments that work for an online community, with a view to refining the technology and methodology for future use within the project and after the completion of the project.

Technology is an aspect that worked exceptionally well in this experiment. Not only did the present researcher have the opportunity of using the functionalities for pre-editing, post-editing and evaluation built during the ACCEPT project, but, had the opportunity of working closely with the in-house developer of the ACCEPT portal, discussing, testing and designing these functionalities making them fit for a real-life application.

This was facilitated by the sponsoring and supportive role provided by a key partner on the project, Symantec, which resulted in privileged access to resources, such the Norton Communities. Third-party technology, such as SurveyMonkey, BLAST and an adapted version of Appraise ran smoothly.

In regards to scheduling, the conducting of the experiments ran smoothly. This included the planning, execution and analysis of a pilot study. More importantly, it involved incorporating the results from the pilot study and modifying the methodology for the full-scale study accordingly. This comprised the preparation, recruitment, the actual post-editing study, planning the evaluation, conducting the evaluation and the analysis of the data. After overcoming initial pre-conceptions about working with online communities, the relationship with the community was initiated by and based on the studies, rather than the Norton products, and the understanding of the community gained was successful and valuable.

Challenges encountered were based on recruitment, firstly, that of the post-editors, i.e. engaging community members to invest about eight hours of their time over a period of several weeks and, secondly, that of evaluators, asking in-house domain specialists to evaluate the post-edited content in addition to their usual work. These challenges were overcome by persistence and frequent personalised communications. A learning point here was that online communities are autonomous and often self-governed and are unpredictable entities that cannot be led but must be gently steered.

The methodological approach taken revealed that the manner of measuring attitudes towards errors and measuring the post-editors' language skills and domain knowledge might not have been fully appropriate for the study. Nonetheless, these were considered

feasible and appropriate methods, based on the context of the experiments. Alternative methods would have required excessively more time and resources and may have resulted in further community attrition. To be able to truthfully measure language skills, a language test in a controlled environment would have been required. Based on the fact that participants were volunteers and recruitment was challenging, this would have potentially prevented participants from taking part. It may also be the case that the post-editing quality depends on aspects of the post-editor profile other than those considered here, such as aspects rooted in the psyche of the post-editors, such as motivation, which has been suggested before, for example by Göpferich (2013).

An aspect of the methodology that is a clear limitation to the study is the scale of the experiment, i.e. 18 as the number of post-editors and the number of segments evaluated (44% of all post-edited content, i.e. 5,100 post-edited segments and 1,275 machine translated segments). Nevertheless, this thesis took a step into proving that volunteer post-editing by lay post-editors is feasible, from which further research can branch out into different directions, such as different domains or language pairs. The research results presented in Chapters 5, 6 and 7 are valid for the community studied here and although they may not be generalisable, i.e. applicable to all online communities, they are transferable, i.e. may provide helpful insight for studies with a similar research design in a comparable context. While the research context was very different to that of previous studies, results suggested similar patterns, such as variance of post-editing quality (e.g. Koehn 2010) or the evaluation of post-editing quality (e.g. Lommel, Popovic and Burchardt 2014). Along with other parallels in themes explored in this concluding chapter, this suggests that the results are indeed transferable.

The error annotation could have been improved by using more than one annotator.

However, a problem with error annotations has proven to be that it is difficult to reach consensus (as extensively explored in Lommel, Popovic and Burchardt (2014), for example). This would be the case here, too, as the post-edited solutions were not compared to the reference translation but the smallest number of errors were tagged that would have needed to be rectified to achieve a correct translation. This evaluation step could have been improved by using more annotators who would be presented with the closest translation to the post-edited solutions as a reference. This would have involved considerably more cost than already generated, such as with the domain specialist evaluation, but would have rendered the evaluation more rigorous and increased its reproducibility.

Methodological aspects that worked exceptionally well were the community evaluation and the evaluation of fluency on a 5-point scale. A first relatively unsuccessful prototype of evaluation designed by the present researcher and described in Mitchell and Roturier (2012) had generated approximately 100 ratings in six months and had set back expectations on feedback from the community. After intensive studying of and engaging with the community and testing the evaluation design in several rounds with subjects unfamiliar with the study, a new design was developed and launched (presented also in Mitchell, Roturier and Silva 2014). Within seconds the first ratings were recorded, with an astonishing 1,470 ratings recorded in a four-week period. Furthermore, the discovery that the evaluation of fluency, which was used both in the domain specialist and the community evaluation was strongly correlated to the TER results (with statistical significance), validated that it was indeed an appropriate measure of determining quality.

All in all, the mixed methods approach to measuring quality worked very well on a

macro level, as it provided insight into quality from three different perspectives and allowed for a differentiated analysis and conclusion. On a micro level, the approach would have benefited from more invasive methods of measuring language and domain competences as described above. A first step to investigating the 'missing links' in the post-editor profile, such as motivation, as identified in this thesis, could be taken by conducting interviews with the participants or by using TAP as a data collection method.

8.5 Recommendations for Future Research

After designing a successful light-weight community evaluation widget, it would be beneficial to expand on it to, firstly, empower the community more as an end-user and secondly to increase the comparability with other evaluation approaches, here the domain specialist evaluation. This could include rating larger chunks of texts, including context or the source and rating aspects other than fluency only, such as adequacy.

Evaluation could also be taken a step further in studying how the end users benefit from the post-edited material by either designing a comprehensibility test, for which the subjects have to answer questions to show that they are able to solve the problem in question, or a practical test for which they actually follow the instructions of a post to solve a problem (see Sinaiko and Brislin 1973). The reason that this has not been implemented yet, is that it requires an increase in resources. For measuring comprehensibility, questions specific to the post-edited content that are of an adequate level of difficulty would need to be designed. Further, subjects would need to be recruited who are at least users of the Norton products and are ideally members of the forum. A practical test would require access to necessary equipment and the ability

of the researchers to reproduce the problem that is described in the content and to have an independent technical expert who would be able to grade the procedure and results.

A concept that has not been studied appropriately here that is likely to influence PE quality is that of user motivation. Investigating motivation and how to engage a community that does not have translation as its main purpose is important as users need to be motivated to engage in post-editing for the concept to be successful. It would need to be studied in combination with other aspects of the user profile, such as language skills in order to identify any possible dependencies in accordance with the Dynamic Systems Theory as pointed out by Göpferich (2013).

8.6   Concluding Remarks

This thesis demonstrated that lay post-editors are able to post-edit user-generated machine translated content in an online community context, albeit with varying degrees of quality, which appears to be largely unrelated to their skill sets and attitudes. It constitutes a first step towards uncovering when community post-editing is successful or unsuccessful. It further provides insight into working with an online community as a platform for post-editing and evaluation. Finally, it challenges the concepts of quality and current evaluation techniques, pointing out bias in low numbers of evaluators and suggesting that evaluators may not need to be linguists or translators to evaluate content reliably.

Bibliography

ACCEPT. 2012. *Data and report from user studies - Year 1.* Available from: `http://www.accept.unige.ch/Products/D8.1_Data_and_report_from_user_studies-Year_1.pdf` [Accessed 22 September 2014].

– 2013a. *Data and report from user studies - Year 2.* Available from: `http://www.accept.unige.ch/Products/D_8_1_2_Data_and_report_from_user_studies_-_Year_2.pdf` [Accessed 22 September 2014].

– 2013b. *Definition of pre-editing rules for English and French.* Available from: `http://www.accept.unige.ch/Products/D2_2_Definition_of_Pre-Editing_Rules_for_English_and_French_with_appendixes.pdf` [Accessed 3 November 2014].

Allen, J. 2003. Post-Editing. *IN: Computers and translation: a translator's guide.* H. Somers (ed.). Amsterdam, Philadelphia: John Benjamins Publishing Company, pp.297–317.

Amazon.com. 2014. *amazonmechanical turk.* Available from: `https://www.mturk.com/mturk/welcome` [Accessed 4 November 2014].

Aranberri, N., Labaka, G., Ilarraze, A. D. de and Sarasola, K. 2014. Comparison of post-editing productivity between professional translators and lay users. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice.* Vancouver, Canada, pp.20–34.

Arnold, D. 1994. *Machine translation: an introductory guide.* Oxford: NCC Blackwell.

Aziz, W., de Sousa, S. C. and Specia, L. 2012. PET: a tool for post-editing and assessing machine translation. *IN: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC), 26 May.* Istanbul, Turkey, pp.3982–3987.

Babych, B., Hartley, A. and Sharoff, S. 2009. Evaluation-guided pre-editing of source text: improving MT-tractability of light verb constructions. *IN: Proceedings of the 13th Annual Conference of the European Association for Machine Translation (EAMT), 14-15 May.* L. Marquez and H. Somers (eds.). Barcelona, Spain, pp.36–42.

Baer, N. 2010. *Trends in crowdsourcing: case studies from not-for-profit and for-profit organisations.* Available from: `http://www.wiki-translation.com/tiki-download_wiki_attachment.php?attId=62` [Accessed 19 September 2014].

Balkan, L., Netterz, K., Arnold, D. and Meijer, S. 1994. Test suites for natural language processing. *IN: Proceedings of Language Engineering Convention, 6-7 July.* Paris, France, pp.17–21.

Banerjee, S. and Lavie, A. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. *IN: Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization held at ACL 2005, 29 June.* Ann Arbour, Michigan, USA, pp.65–72.

Belam, J. 2003. Buying up to falling down. A deductive approach to teaching post-editing. *IN: Proceedings of MT Summit IX Workshop on Teaching Translation Technologies and Tools (T4).* New Orleans, USA, pp.10.

Bernth, A. and Gdaniec, C. 2002. MTranslatability. *Machine Translation*, 16(3), pp.175–218.

Bishop, J. 2007. Increasing participation in online communities: A framework for human-computer interaction. *Computers in Human Behaviour*, 23(4), pp.1881–1893.

Blanchard, A. 2004. Virtual behavior settings: an application of behavior setting theories to virtual communities. *Journal of Computer Mediated Communication [Online]*, 9(2). Available from: `http://jcmc.indiana.edu/vol9/issue2/blanchard.html` [Accessed 20 April 2012].

Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Monz, C., Post, M. and Specia, L. 2014. *Proceedings of the 9th Workshop on Statistical Machine Translation (WMT) held at ACL 2014, 26-27 June.* Baltimore, Maryland, USA: Association for Computational Linguistics.

Bouillon, P., Gaspar, L., Gerlach, J. and Roturier, J. 2014. Pre-editing by forum users: a case study. *IN: Proceedings of the Workshop on Controlled Natural Language Simplifying Language Use (LREC-CNL), 27 May.* Reykjavik, Iceland, pp.3–10.

Bowker, L. and Pearson, J. 2002. *Working with specialized language: a practical guide to using corpora.* London/New York: Routledge.

Callison-Burch, C., Osborne, M. and Koehn, P. 2006. Re-evaluating the role of BLEU in machine translation r esearch. *IN: Proceedings the Eleventh Conference of the European Chapter of the Association for Computational Linguistics (EACL), 3-7 April.* Trento, Italy, pp.249–256.

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C. and Schroeder, J. 2007. (Meta-) evaluation of machine translation. *IN: Proceedings of the 2nd Workshop on Statistical Machine Translation (WMT) held at ACL 2007, 23 June.* Prague, Czech Republic, pp.136–158.

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C. and Schroeder, J. 2008. Further meta-evaluation of machine translation. *IN: Proceedings of the 3rd Workshop on Statistical Machine Translation (WMT) held at HLT 2008, 19 June.* Columbus, Ohio, pp.70–106.

Callison-Burch, C., Koehn, P., Monz, C. and Schroeder, J. 2009. Findings of the 2009 workshop on statistical machine translation. *IN: Proceedings of the 4th Workshop on Statistical Machine Translation (WMT) held at EACL 2009, 30-31 March.* Athens, Greece, pp.1–28.

Callison-Burch, C., Koehn, P., Monz, C. and Zaidan, O. F. 2012a. *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT) held at EMNLP 2011, 30-31 July.* Edinburgh, Scotland: Association for Computational Linguistics.

Callison-Burch, C., Koehn, P., Monz, C., Post, M., Soricut, R. and Specia, L. 2012b. *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT) held at NAACL 2012, 7-8 June.* Montréal, Canada: Association for Computational Linguistics.

CasMaCat. 2013. *Review Presentations.* Available from: `http://www.casmacat.eu/index.php?n=Main.SecondYear` [Accessed 22 September 2014].

Castilho, S., O'Brien, S., Alves, F. and O'Brien, M. 2014. Does post-editing increase usability? A study with Brazilian Portuguese as target language. *IN: Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 16-18 June.* Dubrovnik, Croatia.

Climent, S., Moré, J., Oliver, A., Salvatierra, M., Sànchez, I., Taulé, M. and Vallmanya, L. 2003. Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation.

Journal of Computer-Mediated Communication, 9(1). DOI: `http://jcmc.indiana.edu/vol9/issue1/climent.html`. [Accessed 22 March 2014].

Coughlin, D. 2003. Correlating automated and human assessments of machine translation quality. *IN: Proceedings of the Machine Translation Summit of the International Association for Machine Translation (MT Summit IX), 23-28 September.* New Orleans, USA, pp.63–70.

Creswell, J. W. 2011. Controversies in mixed methods research. *IN: The SAGE Handbook of Qualitative Research.* N. K. Denzin and Y. S. Lincoln (eds.). 4 ed. Thousand Oaks, CA: Sage.

Creswell, J. W. and Plano Clark, V. L. 2007. *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

– 2011. *Designing and conducting mixed methods research.* 2nd ed. Thousand Oaks, CA: Sage.

CrowdFlower. 2014. *CrowdFlower.* Available from: `http://www.crowdflower.com/` [Accessed 4 November 2014].

Cushing, E. 2013. *Amazon Mechanical Turk: the digital sweatshop.* Available from: `http://www.utne.com/science-and-technology/amazon-mechanical-turk-zm0z13jfzlin.aspx#axzz3DfN1z64Y` [Accessed 18 September 2014].

de Almeida, G. 2013. *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages.* PhD thesis. Dublin City University.

de Almeida, G. and O'Brien, S. 2010. Analysing post-editing performance: correlations with years of translation experience. *IN: Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT), 27-28 May.* St. Raphaël, France, pp.8.

Denkowski, M. J. and Lavie, A. 2011. METEOR 1.3: automatic metric for reliable optimization and evaluation of machine translation systems. *IN: Proceedings of the 6th Workshop on Statistical Machine Translation.* Edinburgh, Scotland, pp.85–91.

Denkowski, M. and Lavie, A. 2010. Choosing the right evaluation for machine translation: an examination of annotator and automatic metric performance on human judgment tasks. *IN: Proceedings of the 9th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2010), 31 October - 4 November.* Denver, Colorado, USA, pp.9.

Denkowski, M. and Lavie, A. 2012. TransCenter: web-based translation research suite. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012.* San Diego, California, USA, pp.1.

Denkowski, M., Lavie, A., Lacruz, I. and Dyer, C. 2014. Real time adaptive machine translation for post-editing with cdec and TransCenter. *IN: Proceedings of EACL Workshop on Humans and Computer-assisted Translation, 26-30 April.* Gothenburg, Sweden, pp.6.

Depraetere, I. 2010. What counts as useful advice in a university post-editing training context? Report on a case study. *IN: Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT), 27-28 May.* St. Raphaël, France, pp.7.

Dodd, S. 2011. *Social[ism] media 2.0.* Available from: `http://translorial.com/2011/01/01/crowdsourcing-socialism-media-2-0/` [Accessed 29 August 2014].

Doddington, G. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. *IN: Proceedings of the Second International Conference on Human Language Technology Research (HLT), 24-27 March.* San Diego, California, USA, pp.128–132.

Doherty, S. and O'Brien, S. 2012. A user-based usability assessment of raw machine translated techincal instructions. *IN: Proceedings of the 10th biennial conference of the Association for Machine Translation in the Americas (AMTA-2012), 28 October - 1 November.* San Diego, California, USA, pp.6.

– 2013. Assessing the usability of raw machine translation output: a user-centered study using eye tracking. *International Journal of Human Computer Interaction*, 30(1), pp.40–51.

Doherty, S. 2012. *Investigating the effects of controlled language on the reading and comprehension of machine translated texts: a mixed-methods approach.* PhD thesis. Dublin City University.

Dombek, M. 2014. *A study into the motivations of internet users contributing to translation crowdsourcing: the case of polish facebook user-translators.* PhD thesis. Dublin City University.

Drugan, J. 2011. Translation ethics wikified: how far do professional codes of ethics and practice apply to non-professionally produced translation. *IN:* M. O'Hagan (ed.). Antwerp: Artesis University College, pp.111–131.

Drugan, J. 2013. *Quality in professional translation: assessment and improvement.* London: Bloomsbury.

Désilets, A. and van der Meer, J. 2011. Co-creating a repository of best-practices for collaborative translators. *IN: Linguistica Antverpienisia New Series – Themes in Translation Studies. Translation as a Social Activity, 10/2011.* M. O'Hagan (ed.). Antwerp: Artesis University College, pp.27–45.

Du, N., Wu, B., Pei, X., Wang, B. and Xu, L. 2007. Community detection in large-scale social networks. *IN: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD), 12-15 August.* San Jose, CA, USA, pp.16–25.

Dyson, M. and Hannah, J. 1987. Toward a methodology for the evaluation of machine-assisted translation systems. *Machine Translation*, 2(2-3), pp.163–176.

European Centre for the Development of Vocational Training CEDEFOP. 2011. *Gemeinsamer europäischer Referenzrahmen für Sprachen - Raster zur Selbsteinschätzung.* Available from: `http://europass.cedefop.europa.eu/de/resources/european-language-levels-cefr` [Accessed 3 April 2014].

Federico, M., Bertoldi, N. and Cettolo, M. 2008. IRSTLM: an open source toolkit for handling large scale language models. *IN: Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH), 22-26 September.* Brisbane, Australia, pp.1–8.

Federmann, C. 2010. Appraise: an open-source toolkit for manual phrase-based evaluation of translations. *IN: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), 19-21 May.* Valetta, Malta, pp.1731–1734.

Fiederer, R. and O'Brien, S. 2009. Quality and machine translation: a realistic objective? *The Journal of Specialised Translation*, 11(January), pp.52–74.

Garcia, I. 2010. Is machine translation ready yet? *Target*, 22(1), pp.7–21.

– 2011. Translating by post-editing: is it the way forward? *Machine Translation*, 25(3), pp.217–237.

Gaspari, F., Toral, A., Naskar, S. K., Groves, D. and Way, A. 2014. Perception vs reality: measuring machine translation post-editing productivity. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice.* Vancouver, Canada, pp.60–72.

Gouadec, D. 2007. *Translation as a profession.* Amsterdam: John Benjamins Publishing Company.

Göpferich, S. 2008. *Translationsprozessforschung: Stand - Methoden - Perspektiven.* Tübingen: Narr.

– 2009a. Process research into the development of translation competence: where are we, and where do we need to go? *Across Languages and Cultures*, 10(2), pp.169–191.

– 2009b. Towards a model of translation competence and its acquisition: the longitudinal study TransComp. *IN: Behind the mind: methods, models and results in translation process research. Copenhagen Studies in Language (CSL).* I. M. M. Susanne Göpferich Arnt Lykke Jakobsen (ed.). Copenhagen: Samdundslitteratur, pp.11–37.

– 2013. Translation competence, Explaining development and stagnation from a dynamic systems perspective. *Interdisciplinarity in Translation and Interpreting Process Research; Special issue of Target*, 25(1), pp.61–76.

Graça, J. 2014. *Crowdsourcing in MT, the 9th Machine Translation Marathon, Keynote Speech.* Available from: `http://statmt.org/mtm14/uploads/Main/Crowdsourcing_MTM2014.pdf` [Accessed 12 October 2014].

Greene, J., Caracelli, V. and Graham, W. 1989. Toward a conceptual framework for a mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), pp.255–274.

Guerberof, A. 2009. Productivity and quality in the postediting of outputs from translation memories and machine translation. *Localisation Focus*, 7(1), pp.11–21.

– 2012. *Productivity and quality in the post-editing of outputs from translation memories and machine translation.* PhD thesis. Universitat Rovira.

– 2014. Correlations between productivity and quality when post-editing in a professional context. *Machine Translation; Special Issue on Post-Editing*, 28(3-4), pp.165–186.

Hagel, J. and Armstrong, A. 1997. Net gain: expanding markets through virtual communities. *McKinsey Quarterly*, 1(Winter), pp.141–153.

Hardt, D. and Elming, J. 2010. Incremental re-training for post-editing SMT. *IN: Proceedings of the 10th biennial conference of the Association for Machine Translation in the Americas (AMTA-2012), 28 October - 1 November.* San Diego, California, USA, pp.10.

Harris, B. and Sherwood, B. 1978. Translating as an innate skill. *IN: Language Interpretation and Communication.* H. S. D. Gerver (ed.). New York, pp.155–170.

Harris, B. 1973. La traductologie, la traduction naturelle, la traduction automatique et la sémantique. *IN: Problèmes de sémantique.* J. McA'Nulty (ed.). Montréal: Presses de l'Université du Québec, pp.133–146.

– 1976. *The importance of natural translation.* Available from: `http://www.academia.edu/1406388/The_importance_of_natural_translation` [Accessed 8 October 2014].

Hönig, H. G. 1991. Holmes' 'Mapping Theory' and the landscape of mental translation processes. *IN: Translation Studies: the state of the art. Proceedings from the first James S. Holmes Symposium on Translation Studies.* T. N. K. van Leuven-Zwart (ed.). Amsterdam: Rodopi, pp.77–89.

– 1995. *Konstruktives Übersetzen.* Tübingen: Stauffenburg.

House, J. 1977. *A model for translation quality assessment.* Tübingen: Narr.

– 1997. *Translation quality assessment: a model revisited.* Tübingen: Narr.

– 2015. *Translation quality assessment: past and present.* London, New York: Routledge.

Hovy, E., King, M. and Popescu-Belis, A. 2002. Principles of context-based machine translation evaluation. *Machine Translation*, 17(1), pp.43–75.

Howe, J. 2006. *Crowdsourcing: a definition.* Available from: `http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html` [Accessed 17 September 2014].

Hu, C., Bederson, B. B. and Resnik, P. 2010. Translation by iterative collaboration between monolingual users. *IN: Proceedings of Graphics Interface 2010 (GI '10), 31 May - 2 June.* Ottawa, Ontario, Canada, pp.39–46.

Hu, C., Bederson, B. B., Resnik, P. and Kronrod, Y. 2011. MonoTrans2: a new human computation system to support monolingual translation. *IN: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), 7-12 May.* Vancouver, BC, Canada, pp.1133–1136.

Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics (NAACL-HLT), Proceedings, June 3-8. 2012. Montréal, Canada: The Association for Computational Linguistics. ISBN: 978-1-937284-20-6.

Iriberri, A. and Leroy, G. 2009. A life-cycle perspective on online community success. *ACM Computing Surveys*, 41(2), pp.11:1–11:29.

Jones, Q. and Rafeli, S. 2000. Time to split, virtually: 'discourse architecture' and 'community building' as means to creating vibrant virtual metropolises. *International Journal of Electronic Commerce and Business Media*, 10(4), pp.214–223.

Kelly, N. 2009. Myths about crowdsourced translation. *Multilingual*, (December 2009), pp.62–3.

Kennedy, G. 1998. *An introduction to corpus linguistics*. London/New York: Longman.

King, M., Popescu-Belis, A. and Hovy, E. 2003. FEMTI: Creating and using a framework for MT evaluation. *IN: Proceedings of the Machine Translation Summit of the International Association for Machine Translation (MT Summit IX), 23-28 September.* New Orleans, USA, pp.224–231.

Kirchhoff, K., Capurro, D. and Turner, A. M. 2014. A conjoint analysis framework for evaluating user preferences in machine translation. *Machine Translation*, 28(1), pp.1–17.

Kneser, R. and Ney, H. 1995. Improved backing-off for the m-gram language modeling. *IN: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'95), 8-12 May.* Detroit, Michigan, USA, pp.181–184.

Knight, K. and Chander, I. 1994. Automated postediting of documents. *IN: Proceedings of the 12th National Conference on Artificial Intelligence (AAAI) Volume 1, 31 July - 4 August.* Seattle, Washington, pp.779–784.

Koehn, P. 2009. A web-based interactive computer aided translation tool. *IN: Proceedings of the ACL Interactive Poster and Demonstration Sessions, 3 August.* Suntec, Singapore, pp.17–20.

Koehn, P. and Monz, C. 2006. Shared task: statistical machine translation between European languages. *IN: Proceedings of the 1st Workshop on Statistical Machine Translation (WMT) held at HLT-NAACL 2006, 8-9 June.* New York, USA, pp.102–121.

Koehn, P. 2010. Enabling monolingual translators: post-editing vs. options. *IN: Human Language Technologies: Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2-4 June.* Los Angeles, California: Association for Computational Linguistics, pp.537–545.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. and Herbst, E. 2007. Moses: open source toolkit for statistical machine translation. *IN: ACL 2007: Proceedings of Demo and Poster Sessions*. Prague, Czech Republic, pp.177–180.

Koponen, M. 2010. Assessing machine translation quality with error analysis. *IN: Electronic proceedings of the KäTu symposium on translation and interpreting studies*. Vol. 4. 1/2010, pp.1–12.

– 2011. *Correctness of machine translation: a machine translation postediting task, 3rd MOLTO Project Meeting, Helsinki, Finland*. Available from: `http://www.molto-project.eu/sites/default/files/molto_20110902_mkoponen.pdf` [Accessed 28 April 2012].

– 2012. Comparing human perceptions of post-editing effort with post-editing operations. *IN: Proceedings of the 7th Workshop on Statistical Machine Translation (WMT) held at NAACL 2012, 7-8 June*. Montreal, Canada: ACL, pp.181–190.

– 2013. This translation is not too bad: an analysis of post-editor choices in a machine translation post-editing task. *IN: Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP), 2-6 September*. S. O'Brien, M. Simard and L. Specia (eds.). Nice, France, pp.1–9.

Koponen, M., Aziz, W., Ramos, L. and Specia, L. 2012. Post-editing time as a measure of cognitive effort. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012*, pp.11–20.

Krings, H. P. 2001. *Repairing texts: empirical investigations of machine translation post-editing processes*. Kent: Kent State University Press.

Lacruz, I., Denkowski, M. and Lavie, A. 2014. Cognitive demand and cognitive effort in post-editing. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice*. Vancouver, Canada, pp.73–84.

Lacruz, I., Shreve, G. M. and Angelone, E. 2012. Average pause ratio as an indicator of cognitive effort in post-editing: a case study. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012*. San Diego, California, USA, pp.21–30.

Landis, J. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1), pp.159–174.

Laurenzi, A., Brownstein, M., Turner, A. M. and Kirchhoff, K. 2013. Integrated post-editing and translation management for lay user communities. *IN: Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice, 2-6 September.* S. O'Brien, M. Simard and L. Specia (eds.). Nice, France, pp.27–34.

Lavie, A. and Agarwal, A. 2007. Meteor: an automatic metric for MT evaluation with high levels of correlation with human judgments. *IN: Proceedings of the 2nd Workshop on Statistical Machine Translation (WMT) held at ACL 2007, 23 June.* Prague, Czech Republic: Association for Computational Linguistics, pp.228–231.

LDC. 2005. *Linguistic data annotation specification: assessment of fluency and adequacy in translations. Revision 1.5.* Available from: `http://www.ldc.upenn.edu/Catalog/docs/LDC2003T17/TransAssess02.pdf` [Accessed 3 October 2012].

Lee, F. S. L., Vogel, D. and Limayem, M. 2003. Virtual community informatics: a review and research agenda. *The Journal of Information Technology Theory and Application (JITTA)*, 5(1), pp.47–61.

Lin, D., Murakami, Y., Ishida, T., Murakami, Y. and Tanaka, M. 2010. Composing human and machine translation services: language grid for improving localization processes. *IN: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), 19-21 May.* Valetta, Malta, pp.500–506.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frakowski, D., Terveen, L., Rashid, A., Resnick, P. and Kraut, R. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication [Online]*, 10(4). Available from: `http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00273.x/full` [Accessed 1 December 2014].

LISA. 2009. *LISA QA Model 3.1.* Available from: `http://www.lisa.org/LISA-QA-Model-3-1.124.0.html` [Accessed 20 August 2010].

Lommel, A. 2014. *A new model of translation quality.* Available from: `http://lodel.irevues.inist.fr/tralogy/index.php?id=319` [Accessed 13 October 2014].

Lommel, A., Popovic, M. and Burchardt, A. 2014. Assessing inter-annotator agreement for translation error annotation. *IN: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), 26-31 May.* Reykjavik, Iceland, pp.7.

MacQueen, J. B. 1967. Some methods for classification and analysis of multivariate observations. *IN: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, pp.281–297.

Martzoukos, S. and Monz, C. 2010. The UvA system description for IWSLT. *IN: Proceedings of the 7th International Workshop on Spoken Language Translation, 2-3 December.* Paris, France, pp.205–208.

MateCat. 2013. *Project periodic report.* Available from: `http://www.matecat.com/reports/second-year-project-periodic-report-3/` [Accessed 22 September 2014].

McAuley, J., O'Conner, A. and Lewis, D. 2012. Community management matters: advanced visual analytics for online community managers. *IN: Innovative approaches of data visualization and visual analytics.* W. H. Mao Lin Huang (ed.). Advances in data mining and database management (ADMDM) book series. IGI Global, pp.1–37.

McDonough Dolmaya, J. 2011. Moral ambiguity: some shortcomings of professional codes of ethics for translators. *The Journal of Specialised Translation*, 15(January), pp.28–49.

McDonough Dolmaya, J. 2012. Analyzing the crowdsourcing model and its impact on public perceptions of translation. *The Translator: studies in intercultural communication*, 18(2), pp.167–191.

McEnery, T., Xiao, R. and Tono, Y. 2006. *Corpus-based Language Studies.* London/New York: Routledge.

Melby, A. K., Housley, J., Fields, P. J. and Tuioti, E. 2012. Reliably assessing the quality of post-edited translation based on formalized structured translation specifications. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012.* San Diego, USA, pp.31–40.

Mesipuu, M. 2012. *Translation crowdsourcing and user-translator motivation at Facebook and Skype. Translation Spaces.* Vol. 1. Amsterdam: John Benjamins, pp.33–53.

Mitchell, L., O'Brien, S. and Roturier, J. 2014. Quality evaluation in community post-editing. *Special issue Post-Editing in Machine Translation*, 28(3-4), pp.237–262.

Mitchell, L. and Roturier, J. 2012. Evaluation of machine-translated user generated content: a pilot study based on user ratings. *IN: Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT), 28-30 May.* Trento, Italy, pp.61–64.

Mitchell, L., Roturier, J. and O'Brien, S. 2013. Community-based post-editing of machine-translated content: monolingual vs. bilingual. *IN: MT Summit XIV Workshop on Post-editing Technology and Practice.* Nice, France, pp.35–43.

Mitchell, L., Roturier, J. and Silva, D. 2014. Using the ACCEPT framework to conduct an online community-based translation evaluation study. *IN: Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 16-18 June.* Dubrovnik, Croatia, pp.155–158.

Monasterio, N. A. 2009. *-ing words in RBMT: multilingual evaluation and exploration of pre- and post- processing solutions.* PhD thesis. Dublin City University.

Moorkens, J. 2012. *Measuring consistency in translation memories: a mixed-methods case study.* PhD thesis. Dublin City University.

Moorkens, J. and O'Brien, S. 2013. User attitudes to the post-editing interface. *IN: Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice, 2-6 September.* S. O'Brien, M. Simard and L. Specia (eds.). Nice, France, pp.19–25.

Moorkens, J., O'Brien, S. and Vreeke, J. 2014. Kanjingo – a mobile app for post-editing. *IN: Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 16-18 June.* Dubrovnik, Croatia, pp.5.

Muntes, V. and Paladini, P. 2012. *Crowd Localization: bringing the crowd in the postediting process.* Available from: `http://www.youtube.com/watch?v=1bgIKPULcsc` [Accessed 10 August 2012].

O'Brien, S. 2002. Teaching post-editing: a proposal for course content. *IN: Sixth EAMT Workshop "Teaching machine translation", 14-15 November.* Manchester, England, pp.99–106.

– 2006a. *Machine-translatability and post-editing effort: an empirical study using translog and choice network analysis.* PhD thesis. Dublin City University.

– 2006b. Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures*, 7(1), pp.1–21.

– 2011. Towards predicting post-editing productivity. *Machine Translation*, 25(3), pp.197–215.

O'Brien, S. and Roturier, J. 2007. How portable are controlled language rules: a comparison of two empirical MT studies. *IN: Proceedings of the 11th Machine Translation Summit of the International Association for Machine Translation (MT Summit XI), 10-14 September.* Copenhagen, Denmark, pp.8.

O'Brien, S., Balling, L. W., Carl, M., Simard, M. and Specia, L. 2014. *Post-editing of machine translation: process and applications.* Newcastle upon Tyne: Cambridge Scholars Publishing.

O'Curran, E. 2014. Translation quality in post-edited versus human-translated segments: a case study. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice.* Vancouver, Canada, pp.113–118.

O'Hagan, M. 2011. Community translation: translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond. *IN: Linguistical Antverpiensa New Series - Themes in Translation Studies. Translation as a Social Activity, 10/2011.* M. O'Hagan (ed.), pp.111–128.

Olohan, M. and Baker, M. 2000. Reporting that in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp.141–158.

PACTE. 2002. Exploratory texts in a study of translation competence. *Conference interpretation and translation*, 4(4), pp.41–69.

– 2003. Building a translation competence model. *IN: Triangulating Translation.* F. Alves (ed.). Amsterdam/Philadelphia: John Benjamins, pp.43–66.

– 2005. Investigating translation competence: conceptual and methodological issues. *Meta*, 50(2), pp.609–619.

– 2009. Results of the validation of the pacte translation competence model: acceptability and decision making. *Across Languages and Cultures*, 10(2), pp.207–230.

– 2011. Results of the validation of the PACTE translation competence model: translation project and dynamic translation index. *IN: IATIS Yearbook 2010.* S. O'Brien (ed.). London: Continuum, pp.33.

Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J. 2002. BLEU: a method for automatic evaluation of machine translation. *IN: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02), 7-12 July.* Philadelphia, USA, pp.311–318.

Parton, K. and McKeown, K. 2010. MT error detection for cross-lingual question answering. *IN: Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), 23 -27 August.* Beijing, China, pp.946–954.

Parton, K., McKeown, K., Iglesias, G. and Gispert, A. 2012. Can automatic post-editing make MT more meaningful? *IN: Proceedings of the 16th Annual Conference*

of the European Association for Machine Translation (EAMT), 28-30 May. Trento, Italy, pp.111–118.

Perrino, S. 2009. User-generated translation: the future of translation in a Web 2.0 environment. *The Journal of Specialised Translation*, 12(July), pp.55–78.

Pierce, J., Carroll, J., Hamp, E., Hays, D., Hockett, A., Oettinger, A. and Perlis, A. 1966. *Language and machines: computers in translation and linguistics. Technical report, Automatic Language Processing Advisory Committee (ALPAC).* Washington DC, USA: Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council, pp.0–187.

Plitt, M. and Masselot, F. 2010. A productivity test of statistical machine translation post-editing in a typical localisation environment. *The Prague Bulletin of Mathematical Linguistics*, 93(January), pp.7–16.

Pym, A. 2003. Redefining translation competence in an electronic age. In defence of a minimalist approach. *Meta*, XLVIII(4), pp.481–497.

– 2011. Translation research terms: a tentative glossary for moments of perplexity and dispute. *IN: From translation research projects 3.* A. Pym (ed.). Tarragona: Intercultural Studies Group, pp.75–110.

– 2012. *Translation skill-sets in a machine-translation age.* Available from: `usuaris.tinet.cat/apym/on-line/2012_competence_pym.pdf` [Accessed 22 September 2014].

QT Launchpad. 2014. *QT Launchpad.* Available from: `http://www.qt21.eu/launchpad/content/about`.

Ray, R. 2009. *Crowdsourcing: the crowd wants to help you reach new markets.* Romainmôtier: Localization Industry Standards Association (LISA).

Reid, M. and Gray, C. 2007. *Online social networks, virtual communities, enterprises, and information professionals - part 1. past and present.* Available from: `http://www.infotoday.com/searcher/jul07/reid_grey.shtml` [Accessed 18 January 2013].

Rheingold, H. 2000. *The virtual community: homesteading on the electronic frontier.* revised. Cambridge: MIT Press.

Rickard, J. 2009. *Translation in the community.* Available from: `htpp://www.localisation.ie/resources/conferences/2009/presentations/LRC_L10N_in_the_Cloud.pdf` [Accessed 17 January 2013].

Rico, C. and Ariano, M. 2014. Defining language dependend post-editing guidelines: the case of the language pair English-Spanish. *IN: Post-editing of machine translation: processes and applications.* S. O'Brien, L. Winther Balling, M. Carl, M. Simard and L. Specia (eds.). Newcastle upon Tyne: Cambridge Scholars Publishing, pp.300–322.

Roturier, J. 2006. *An investigation into the impact of controlled English rules on the comprehensibility, usefulness, and acceptability of machine-translated technical documentation for French and German users.* PhD thesis. Dublin City University.

Roturier, J. and Bensadoun, A. 2011. Evaluation of MT Systems to Translate User Generated Content. *IN: Proceedings of the 13th Machine Translation Summit of the International Association for Machine Translation (MT Summit XIII), 19-23 September.* Xiamen, China, pp.244–251.

Roturier, J., Mitchell, L. and Silva, D. 2013. The ACCEPT Post-Editing environment: a flexible and customisable online tool to perform and analyse machine translation post-editing. *IN: MT Summit XIV Workshop on Post-editing Technology and Practice.* Nice, France, pp.119–128.

SAE-International. 2014. *SAE J2450 Translation Quality Metric Task Force: Quality Metric for Language Translation of Service Information.* Available from: `http://www.sae.org/standardsdev/j2450p1.htm` [Accessed 4 November 2014].

Sanchis-Trilles, G., Alabau, V., Buck, C., Germann, U., Carl, M., Casacuberta, F., García-Martínez, M., González-Rubio, J., Hill, R., Koehn, P., Leiva, L., Mesa-Lao, B. and Daniel Ortiz-Martínez Hervé Saint-Amand, C. T. 2014. Interactive translation prediction vs. conventional post-editing in practice: a study with the Casmacat Workbench. *Machine Translation; Special Issue on Post-Editing*, 28(3-4), pp.217–235.

Schlosser, A. 2005. Posting versus lurking: communicating in a multiple audience context. *Journal of Consumer Research*, 32(2), pp.260–265.

Schwartz, L. 2014. Monolingual post-editing by a domain expert is highly effective for translation triage. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice.* Vancouver, Canada, pp.34–44.

Seretan, V., Bouillon, P. and Gerlach, J. 2014. A large-scale evaluation of pre-editing strategies for improving user-generated content translation. *IN: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), 26-31 May.* Reykjavik, Iceland, pp.1793–1799.

Sinaiko, H. and Brislin, R. 1973. Evaluating language translations: experiments on three assessment methods. *Journal of Applied Psychology*, 57(3), pp.328–334.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Weischedel, R. 2006. A study of translation error rate with targeted human annotation. *IN: Proceedings of the 7th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2006), 8-12 August.* Cambridge, Massachusetts, USA, pp.223–231.

StataCorp. 2014. *kappa - Interrater agreement.* Available from: `http://www.stata.com/manuals13/rkappa.pdf` [Accessed 12 February 2014].

Stymne, S. 2011. BLAST: a tool for error analysis of machine translation output. *IN: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Student Session (ACL-HLT), 19-24 June.* Portland, Oregon, pp.56–61.

Tashakkori, A and Teddlie, C. 1998. *Mixed methodology: combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Tatsumi, M. 2009. Correlation between automatic evaluation metric scores, post-editing speed, and some other factors. Ottawa, Ontario, Canada, pp.332–339.

– 2010. *Post-editing machine translated text in a commercial setting: observation and analysis.* PhD thesis. Dublin City University.

Tatsumi, M., Aikawa, T., Yamamoto, K. and Isahara, H. 2012. How Good Is Crowd Post-Editing? Its Potential and Limitations. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012.* San Diego, USA, pp.69–77.

TAUS. 2010. *MT post-editing guidelines.* Available from: `https://www.taus.net/postediting/machine-translation-post-editing-guidelines` [Accessed 20 April 2013].

– 2013a. *DQF - the industry's benchmark for translation quality evaluation.* Available from: `https://evaluation.taus.net/about` [Accessed 18 September 2014].

– 2013b. *Knowledgebase.* Available from: `https://evaluation.taus.net/knowledgebase` [Accessed 18 September 2014].

– 2014. *Post-editing course.* Available from: `https://evaluation.taus.net/post-editing-course-pricing` [Accessed 4 November 2014].

Taylor, K. B. and White, J. S. 1998. Predicting what MT is good for: user judgements and task performance. *IN: Machine Translation and the Information Soup.* D. Farwell, L. Gerber and E. H. Hovy (eds.). Berlin: Springer, pp.364–373.

Teixeira, C. S. C. 2014. Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. *IN: Proceedings of the Third Workshop on Post-editing Technology and Practice.* Vancouver, Canada, pp.45–59.

Thelen, E. and Smith, L. B. 1994. *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Toury, G. 1980. *In search of a theory of translation.* Tel Aviv, Israel: The Porter Institute for Poetics and Semiotics.

– 1995. *Descriptive translation studies and beyond.* Amsterdam: John Benjamins Publishing Company.

– 2012. *Descriptive translation studies.* 2nd ed. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Turian, J., Shen, L. and Melamed, D. 2003. Evaluation of machine translation and its evaluation. *IN: Proceedings of the Machine Translation Summit of the International Association for Machine Translation (MT Summit IX), 23-28 September.* New Orleans, USA, pp.386–393.

van Gelder, T. 1998. The dynamical hypothesis in cognitive science. *Behavioural and brain sciences,* 21(5), pp.615–628.

van Slype, G. 1979. *Critical study of methods for evaluating the quality of machine translation. Technical Report BR19142.* Available from: `http://ww.issco.unige.ch/projects/isle/van-slype.pdf` [Accessed 20 April 2012].

Vanderwende, L., Daume III, H. and Kirchhoff, K., eds. 2013. *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics (NAACL-HLT), Proceedings, June 9-14.* Atlanta, Georgia, USA: The Association for Computational Linguistics.

Vieira, L. 2014. Indices of Cognitive Effort in Machine Translation Post-Editing. *Machine Translation; Special Issue on Post-Editing,* 28(3-4), pp.187–216.

Vieira, L. and Specia, L. 2011. A review of machine translation tools from a post-editing perspective. *IN: 3rd join EM+CNGL workshop bringing MT to the user: research emets translators (JEC 2011), 14 October.* Luxembourg, pp.33–42.

Vilar, D., Xu, J., D'Haro, L. and Ney, H. 2006. Error analysis of statistical machine translation output. *IN: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC), 22-28 May.* Genoa, Italy, pp.697–702.

Volk, M. 2008. The automatic translation of film subtitles. A machine translation success story? *IN: Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein.* D. Nivre J. and B. Megyesi (eds.). Sweden: Uppsala University, pp.202–214.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M. and Haythornthwaite, C. 1996. Computer networks as social networks: collaborative work, telework and virtual community. *Annual Reviews Sociology*, 22(1), pp.213–238.

Welocalize and CNGL. 2014. *Translation Productivity*. Available from: `http://www.cngl.ie/wp-content/uploads/2014/03/Translation-Productivity_v4.pdf` [Accessed 13 October 2014].

Wenger, E., McDermott, R. and Snyder, W. 2002. *Cultivating communities of practice: a guide to managing knowledge.* Cambridge, Massachussets: Harvard Business School Press.

Yamashita, N. and Ishida, T. 2006. Effects of machine translation on collaborative work. *IN: Proceedings of the 20th anniversary conference on Computer supported cooperative work (CSCW), 4-8 November.* Banff, Alberta, Canada, pp.515–524.

Yan, R., Gao, M., Pavlick, E. and Callison-Burch, C. 2014. Are two heads better than one? Crowdsourced translation via a two-step collaboration of non-professional translators and editors. *IN: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers, 26-27 June.* Baltimore, Maryland, pp.1134–1144.

Zhechev, V. 2012. Machine Translation Infrastructure and Post-editing Performance at Autodesk. *IN: Proceedings of the Second Workshop on Post-editing Technology and Practice (WPTP 2012) held at AMTA 2012.* San Diego, California, USA, pp.87–96.

Appendix A

Ethical Approval

A-2

Ollscoil Chathair Bhaile Átha Cliath
Dublin City University

DCU

Ms. Linda Mitchell
SALIS

23rd July 2012

REC Reference: DCUREC/2012/129

Proposal Title: **Monolingual post-editing in an online community - An
                         Investigation into motivation and feasibility**

Applicants: Ms. Linda Mitchell, Dr. Sharon O'Brien, Dr. Fred Hollowood,
                         Dr. Johann Roturier

Dear Linda,

Further to expedited review, the DCU Research Ethics Committee approves this
research proposal. Materials used to recruit participants should note that ethical
approval for this project has been obtained from the Dublin City University Research
Ethics Committee. Should substantial modifications to the research protocol be
required at a later stage, a further submission should be made to the REC.

Yours sincerely,

Dónal O'Mathúna

Dr. Donal O'Mathuna
Chairperson
DCU Research Ethics Committee

DCU Research & Innovation

Appendix B

Pre-Task Survey

Umfrage

Einverständniserklärung

In diesem Teil der Umfrage, werden die Studie und ihre Aufgaben erklärt. Ihr Einverständnis zu allen Teilen ist erforderlich, damit Sie an dieser Studie teilnehmen können.

Teilnahmeinformationen

Der Titel der Studie ist „Gemeinschaftsbasiertes Nachbearbeiten maschinell übersetzter Inhalte" („Community-based post-editing of machine-translated content"). Diese Studie wird im Rahmen des Fachbereiches der Angewandten Sprachen und Interkulturellen Wissenschaften (School of Applied Languages and Intercultural Studies) in der Fakultät der Geisteswissenschaften und Sozialwissenschaften durchgeführt. Linda Mitchell ist die leitende Forschungswissenschaftlerin dieser Studie, die über Email linda.mitchell7@mail.dcu.ie oder Telefon +353858122880 erreicht werden kann. Weitere an diesem Experiment beteiligte Wissenschaftler sind, Dr. Sharon O'Brien, Dozentin an der Dublin City University, Dr. Fred Hollowood, Leiter der sprachtechnischen Forschung bei Symantec und Dr. Johann Roturier, leitender Forschungsingenieur bei Symantec. Diese Forschungsstudie wird als Teil des ACCEPT-Projekts (Bewillignungsnummer: 288769) von der EU-Kommission subventioniert.

Sie werden in dieser Studie über einen Zeitraum von zwei Wochen zwölf Texte nachbearbeiten, die automatisch vom Englischen ins Deutsche übersetzt wurden. Davor und danach werden Sie gebeten eine Umfrage zu Ihren Einstellungen und zu Ihren Empfindungen, die Sie während der Studie hatten, auszufüllen. Insgesamt sollte dies nicht mehr als drei Stunden in Anspruch nehmen.

Die Studie ermöglicht es Ihnen, Fähigkeiten zu entwickeln oder zu verbessern um maschinell übersetzte Texte nachzubearbeiten. Des Weiteren wird dies der Norton Community langfristig gesehen zugute kommen, da sich die Möglichkeit bietet, die Inhalte der Community zu bereichern und schnelleren Zugriff zu Lösungsansätzen zu erlangen. Die Ergebnisse dieser Studie werden nach Vollendung der Doktorarbeit (Ende 2014) auf folgender Website öffentlich zugänglich sein: http://doras.dcu.ie/.

Die im Zuge der Studie erfassten Daten werden entsprechend den Vorschriften, die von der EU-Kommission für Forschungsprojekte aufgesetzt wurden, gesammelt, gespeichert und vernichtet. Die Teilnahme an dieser Forschungsstudie ist freiwillig. Jeder Teilnehmer kann die Studie zu jeder Zeit ohne negative Folgen, abbrechen.

Falls Teilnehmer Bedenken bezüglich dieser Studie haben sollten und Kontakt zu einer unabhängigen Person aufnehmen wollen, kontaktieren Sie bitte:
The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel +35317008000

1. Ich habe die zusammenfassende Erklärung gelesen.
2. Ich verstehe die Informationen, die mir präsentiert wurden.
3. Ich hatte die Möglichkeit Fragen zu stellen und diese Studie zu besprechen.
4. Ich habe befriedigende Antworten auf alle meine Fragen erhalten.
5. Ich bin mir darüber im Klaren, dass ich Nachbearbeitungsaufgaben durchführen werde.
6. Ich bin mir darüber im Klaren, dass ich zwei Umfragen ausfüllen werde.
7. Ich bin mir darüber im Klaren, dass die nachbearbeiteten Texte im Rahmen des ACCEPT-Projekts benutzt werden.

1. Ich stimme den oben aufgeführten Aussagen zu:

○ ja

○ nein

Umfrage

B-3

Einverständniserklärung

Meine Teilnahme an dieser Studie ist freiwillig. Ich kann die Forschungsstudie zu jeder Zeit und ohne jegliche negative Folgen abbrechen. Die Daten, die mit meiner Hilfe in dieser Studie erfasst werden, werden entsprechend der Vorschriften, die von der EU-Kommission für Forschungsprojekte aufgesetzt wurden, erfasst, gespeichert und vernichtet. Ich bin mir darüber im Klaren, dass ich auf die Ergebnisse dieser Studie auf http://doras.dcu.ie/ nach Vollendung der Doktorarbeit (Ende 2014) zugreifen kann.

Ich habe die Informationen gelesen und verstanden. Meine Fragen und Bedenken wurden von den Forschern beantwortet, und ich habe eine Kopie dieser Erklärung erhalten.

2. Ich stimme den Bedingungen dieser Studie zu:

○ ja

○ nein

Umfrage

Allgemein

Hier möchten wir Sie bitten ein paar allgemeine Fragen zu beantworten.

3. Bitte geben Sie Ihr Geschlecht an:

○ weiblich

○ männlich

○ Keine Angabe

4. Bitte geben Sie Ihre Alterskategorie an:

○ unter 18

○ 18 - 24

○ 25 - 34

○ 35 - 44

○ 45 - 54

○ 55 - 64

○ über 65

○ keine Angabe

5. Bitte stufen Sie Ihre *Englischkenntnisse* (Lesen) ein. Markieren Sie die Kategorie zu der Sie sich am ehesten zugehörig fühlen.

○ A1 Ich kann einzelne vertraute Namen, Wörter und ganz einfache Sätze verstehen, z. B. auf Schildern, Plakaten oder in Katalogen.

○ A2 Ich kann ganz kurze, einfache Texte lesen. Ich kann in einfachen Alltagstexten (z. B. Anzeigen, Prospekten, Speisekarten oder Fahrplänen) konkrete, vorhersehbare Informationen auffinden und ich kann kurze, einfache persönliche Briefe verstehen.

○ B1 Ich kann Texte verstehen, in denen vor allem sehr gebräuchliche Alltags oder Berufssprache vorkommt. Ich kann private Briefe verstehen, in denen von Ereignissen, Gefühlen und Wünschen berichtet wird.

○ B2 Ich kann Artikel und Berichte über Probleme der Gegenwart lesen und verstehen, in denen die Schreibenden eine bestimmte Haltung oder einen bestimmten Standpunkt vertreten. Ich kann zeitgenössische literarische Prosatexte verstehen.

○ C1 Ich kann lange, komplexe Sachtexte und literarische Texte verstehen und Stilunterschiede wahrnehmen. Ich kann Fachartikel und längere technische Anleitungen verstehen, auch wenn sie nicht in meinem Fachgebiet liegen.

○ C2 Ich kann praktisch jede Art von geschriebenen Texten mühelos lesen, auch wenn sie abstrakt oder inhaltlich und sprachlich komplex sind, z. B. Handbücher, Fachartikel und literarische Werke.

Kommentar

Umfrage

6. Bitte stufen Sie Ihre *Deutschkenntnisse* (Schreiben) ein. Markieren Sie die Kategorie zu der Sie sich am ehesten zugehörig fühlen.

○ A1 Ich kann eine kurze einfache Postkarte schreiben, z. B. Feriengrüsse. Ich kann auf Formularen, z. B. in Hotels, Namen, Adresse, Nationalität usw. eintragen.

○ A2 Ich kann kurze, einfache Notizen und Mitteilungen schreiben. Ich kann einen ganz einfachen persönlichen Brief schreiben, z. B. um mich für etwas zu bedanken.

○ B1 Ich kann über Themen, die mir vertraut sind oder mich persönlich interessieren, einfache zusammenhängende Texte schreiben. Ich kann persönliche Briefe schreiben und darin von Erfahrungen und Eindrücken berichten.

○ B2 Ich kann über eine Vielzahl von Themen, die mich interessieren, klare und detaillierte Texte schreiben. Ich kann in einem Aufsatz oder Bericht Informationen wiedergeben oder Argumente und Gegenargumente für oder gegen einen bestimmten Standpunkt darlegen. Ich kann Briefe schreiben und darin die persönliche Bedeutung von Ereignissen und Erfahrungen deutlich machen.

○ C1 Ich kann mich schriftlich klar und gut strukturiert ausdrücken und meine Ansicht ausführlich darstellen. Ich kann in Briefen, Aufsätzen oder Berichten über komplexe Sachverhalte schreiben und die für mich wesentlichen Aspekte hervorheben. Ich kann in meinen schriftlichen Texten den Stil wählen, der für die jeweiligen Leser angemessen ist.

○ C2 Ich kann klar, flüssig und stilistisch dem jeweiligen Zweck angemessen schreiben. Ich kann anspruchsvolle Briefe und komplexe Berichte oder Artikel verfassen, die einen Sachverhalt gut strukturiert darstellen und so dem Leser helfen, wichtige Punkte zu erkennen und sich diese zu merken. Ich kann Fachtexte und literarische Werke schriftlich zusammenfassen und besprechen.

Kommentar

Umfrage

Norton Forum

Hier möchten wir Sie bitten Fragen zu Ihrer Erfahrung mit Norton und dem Norton Forum zu beantworten.

7. Wie würden Sie die Wichtigkeit der folgenden Aspekte im Norton Forum für sich persönlich auf einer Skala von 1 (unwichtig) bis 5 (wichtig) einstufen?

	1 (unwichtig)	2 (eher unwichtig)	3 (weder noch)	4 (eher wichtig)	5 (wichtig)
Inhalt	○	○	○	○	○
Rechtschreibung	○	○	○	○	○
Grammatik	○	○	○	○	○
Zeichensetzung	○	○	○	○	○

Umfrage

B-7

Projektspezifisch

Hier möchten wir Sie bitten, Fragen speziell zum Projekt zu beantworten.

8. Bitte geben Sie Ihre Meinung zu den folgenden Aussagen auf einer Skala von 1 bis 5 an.

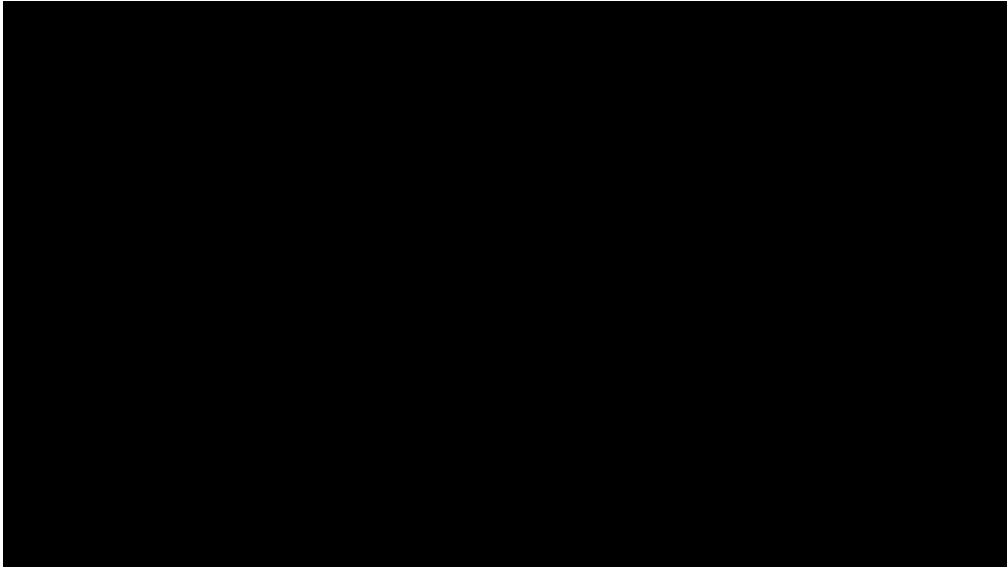|	1 (stimme nicht zu)	2 (stimme eher nicht zu)	3 (weder noch)	4 (stimme eher zu)	5 (stimme zu)
Maschinelle Übersetzung kann im Norton Forum nützlich sein.	○	○	○	○	○
Ich bin motiviert maschinell übersetzte Beiträge zu bearbeiten.	○	○	○	○	○

17. Bitte sehen Sie das Trainingsvideo an, in welchem erklärt wird, wie der Editor funktioniert. Vielen Dank!

Survey

Informed Consent

This part of the survey presents the study and the tasks involved. Your consent is required for all parts in order for you to take part in the study.

Plain Language Statement

The title of this study is "Monolingual post-editing in an online community - an investigation into motivation and feasibility". This experiment is conducted within the School of Applied Languages and Intercultural Studies (SALIS) in the faculty of humanities and social sciences. The principal investigator is Linda Mitchell. Other investigators involved in this experiment are Dr. Sharon O'Brien, senior lecturer at DCU, Dr. Fred Hollowood, Director for Language Technologies Research in Symantec and Dr. Johann Roturier, Principal Research Engineer in Symantec. This research study is funded by the European Commission as part of the ACCEPT project (Grant Agreement Number: 288769).

Your role in this research study involves editing twelve texts that have been automatically translated from English into German over a duration of two weeks. Before and after editing you will be asked to complete a survey on your attitudes and sentiments during the study. All in all, this should not take longer than three hours.

Taking part in this research study will provide you with the opportunity to develop or improve post-editing skills of machine translated output. Furthermore, it will help the Norton Community in the long run, as the forum content will be enriched and faster access to solutions will be provided. The results of this study will be publicly available on http://doras.dcu.ie, once the PhD thesis is completed (late 2014).

The data that is collected during this study will be collected, stored and destroyed in compliance with the regulations for research projects set up by the European Commission. All participants will be assigned anonymous identifiers to preserve anonymity. Involvement in this research study is voluntary. Participants may withdraw from this research study at any point without any negative consequences.

If participants have concerns about this study and wish to contact an independent person, please contact:
The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel 01-7008000

1. I have read the Plain Language Statement (or had it read to me)
2. I understand the information provided
3. I have had an opportunity to ask questions and discuss this study
4. I have received satisfactory answers to all my questions
5. I am aware that I will be required to perform post-editing tasks
6. I am aware that I will be required to complete two surveys
7. I am aware that the post-edited data will be used within the framework of the ACCEPT project

1. I agree with the statements presented above:

○ yes

○ no

Informed Consent

My involvement in this study is voluntary. I may withdraw from the Research Study at any point without any negative consequences. The data that is collected with my help during this study will be collected, stored and destroyed in compliance with the regulations for research projects set up by the European Commission. I am aware that I can access the results of this study on http://doras.dcu.ie/, once the PhD thesis is completed (late 2014).

II have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form.

Survey

2. Therefore, I consent to take part in this research project:

○ yes

○ no

Survey

General

Here we would like to ask you to answer a few general questions.

3. Please indicate your gender:

○ female

○ male

○ I don't wish to specify

4. Please indicate your age group

○ under 18

○ 18 - 24

○ 25 - 34

○ 35 - 44

○ 45 - 54

○ 55 - 64

○ over 65

○ I don't wish to specify

5. Please rate your *English skills* (reading). Indicate which category you feel you belong to the most.

○ A1 I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.

○ A2 I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.

○ B1 I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.

○ B2 I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.

○ C1 I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.

○ C2 I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

Kommentar

[   ]

Survey

6. Please rate your *German skills* (writing). Indicate which category you feel you belong to the most.

○ A1 I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.

○ A2 I can write short, simple notes and messages. I can write a very simple personal letter, for example thanking someone for something.

○ B1 I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

○ B2 I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.

○ C1 I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select a style appropriate to the reader in mind.

○ C2 I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Kommentar

Survey   B-12

Norton Community

Here, we would like you to answer questions on your experience in the Norton Community.

7. How would you rate the importance of the following aspects in the Norton Forum for yourself on a scale of 1 (not important) to 5 (important)?

	1 (not important)	2 (rather not important)	3 (neither nor)	4 (rather important)	5 (important)
grammar	○	○	○	○	○
content	○	○	○	○	○
spelling	○	○	○	○	○
punctuation	○	○	○	○	○

Survey
B-13

Project-related

Here, we would like you to answer questions related to the project itself.

8. Please indicate your opinion on the following statements on a scale of 1 to 5.

	1 (don't agree)	2 (rather don't agree)	3 (neither nor)	4 (rather agree)	5 (agree)
Machine Translation can be useful in the Norton Forum.	○	○	○	○	○
I am motivated to correct machine translated posts.	○	○	○	○	○

17. Please watch the training video, which explains how the editor works. Thank you very much!
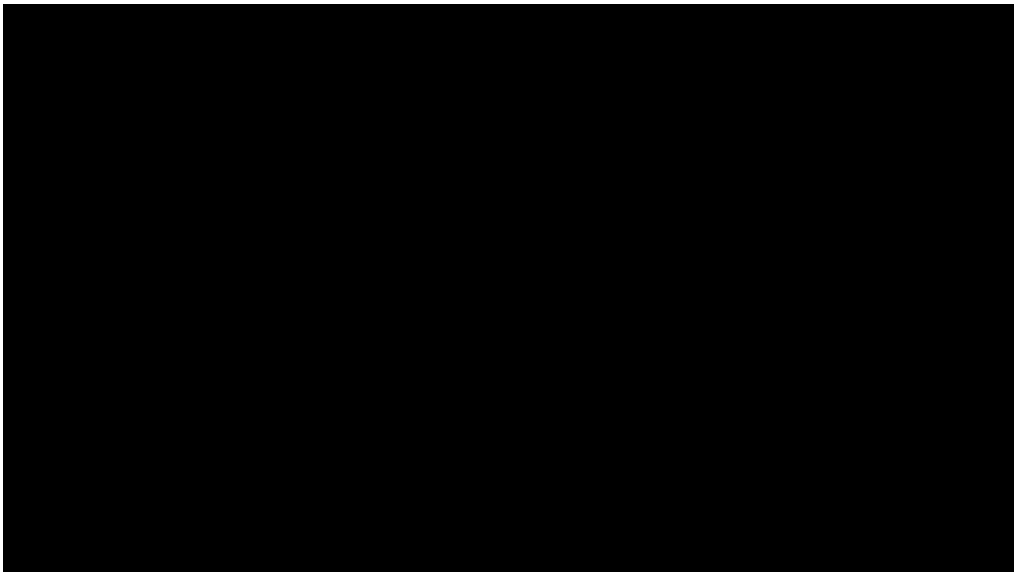
Appendix C

Post-Task Survey

The post-task survey was completed in German by the participants. The German survey is presented first, followed by an English translation of the survey.

Abschlussumfrage
C-2

Meinung nach der Studie

Hier möchten wir Sie bitten, Fragen zum Projekt zu beantworten.

1. Wie würden Sie die Wichtigkeit der folgenden Aspekte im Norton Forum für sich persönlich auf einer Skala von 1 (unwichtig) bis 5 (wichtig) einstufen?

	1 (unwichtig)	2 (eher unwichtig)	3 (weder noch)	4 (eher wichtig)	5 (wichtig)
Zeichensetzung	○	○	○	○	○
Rechtschreibung	○	○	○	○	○
Grammatik	○	○	○	○	○
Inhalt	○	○	○	○	○

2. Bitte geben Sie Ihre Meinung zu den folgenden Aussagen auf einer Skala von 1 bis 5 an.

	1 (stimme nicht zu)	2 (stimme eher nicht zu)	3 (weder noch)	4 (stimme eher zu)	5 (stimme zu)
Maschinelle Übersetzung kann im Norton Forum nützlich sein.	○	○	○	○	○
Ich bin motiviert maschinell übersetzte Beiträge zu bearbeiten.	○	○	○	○	○

Final Survey

C-3

Sentiments after the study

Here, we would like you to answer questions on the project.

1. How would you rate the importance of the following aspects in the Norton Forum for yourself on a scale of 1 (not important) to 5 (important)?

	1 (not important)	2 (rather not important)	3 (neither nor)	4 (rather important)	5 (important)
grammar	○	○	○	○	○
content	○	○	○	○	○
spelling	○	○	○	○	○
punctuation	○	○	○	○	○

2. Please indicate your opinion on the following statements on a scale of 1 to 5.

	1 (don't agree)	2 (rather don't agree)	3 (neither nor)	4 (rather agree)	5 (agree)
Machine Translation can be useful in the Norton Forum.	○	○	○	○	○
I am motivated to correct machine translated posts.	○	○	○	○	○

Appendix D

Source Texts - Group A and B

Task A1 - English ST:

W32.Downadup.B

Hi, My Mac has been infected with W32.Downadup.B. Every resource says that this infection is only for PCs and offers solutions to resolve it on PCs. I need to remove it, because it is slowing down my browser and other computers on my network. I really just want to watch netflix without ridiculous load times.

I think on one attempt you successfully removed the file (running the rm command after success would show "file not found") Hopefully this will prevent the repeated reports, but if not, let me know.

Task A2 - English ST:

Copy to clip board not working for red threats detected screen

In the past, using Norton NIS 2012 when I copied to clip board the tracking cookies would be listed when I pasted them to a notepad. Now using the new version after re- subscribing which doesn't say NIS 2013 it only says (Norton Internet Security) ver. 20.2.1.22. The tracking cookie information is omitted from the paste. I use this information to block stuff in my IE9 browser privacy sites. Is this normal now with the ver. 20.2.1.22 ? I have tried uninstalling with the start menu and after that using the

Norton_Removal_Tool as well. After reinstalling everything works the same, no specific tracking cookie information on the paste. No, it's not normal; it's a defect present in the new Norton 2013 products. The 'Show' dropdown box in all of the 'File Insight' windows has been redesigned to support the Windows 8 Touch capability.

As an unintended consequence of this design change, the "Copy to Clipboard' feature present in this dialog box no longer functions as expected. To work around the issue, you need to run a Quick Scan. Then select the 'Export Results' link shown below to export the scan results to a text file: Open this newly saved text file. You should now see the detailed tracking cookie information. This will assist you with your decision as to whether or not you should block each of the sites listed. Hope this helps.

Task A3 - English ST:

File Guard won't open files even with correct administrator password.

The title pretty much says it all. I have to turn file guard off to get to my protected files because it won't open even with the correct administrator name and password. This started happening a couple of days ago and it worked fine before then. Any ideas?

Hi, this may be a problem with File Guard populating the user name field with the full name instead of the short (aka, posix) user name. To determine the short user name: 1. Open /Applications/Utilities/Terminal.app 2. Enter: whoami 3. Press return. Your short (posix) user name will be shown by the whoami command. Use that name when File Guard prompts for your admin password. Please let us know if this helps. Thanks!

Task A4 - English ST:

Worried about hidden virus on NIS 20.3.0.36

Two days ago, stupid me clicked on a link contained in an Email I thought was from a friend. As soon as it opened a page, I realized the Email had not been sent by a friend.... She had been hacked. I have done three scans with NIS and a full scan with free Malwarebytes Anti-Malware. All came out with clean results, and my 64-bit Windows 7 computer is acting normal. Is there anything further I should do to make

sure a hacker has not dropped a file on my system? I am hesitant to use the Norton Power Eraser Tool. My thanks....

Hi, entries in Norton Product Tamper Protection record any attempt by an outside agent to access a Norton file or process. In almost all cases, the events are caused by legitimate programs or Windows processes, and are nothing to worry about." That is true of the entries you have posted. Those are all legitimate things running on your PC. Norton is simply making sure that nothing, not even Windows, can cause problems by disrupting Norton's operation. There is no evidence of anything malicious in what you have posted.

Task A5 - English ST:

NEED HELP FAST!!

When I do ctrl ALT delete and click start task manager and then go under processes. What is Norton called? I need to know asap please!! !

Hi, are you sure Norton is blocking Teamviewer? Open the Program Control panel in Norton (Settings > Network >Smart Firewall > Program Control). If you see Teamviewer listed there as "Auto," then Norton is not blocking it and the problem lies elsewhere. You can, however, remove the Teamviewer entry from the list to automatically repair any possible misconfiguration. First, make sure Teamviewer is not running. Then click on the Teamviewer entry and click Remove, Apply and OK. Exit out of Norton and restart Teamviewer. Norton will automatically configure new rules for the program and it should work. You cannot kill the Norton process. It is hardened against outside tampering.

Task A6 - English ST: What happens when a NAV license expires?

Is it only unable to update (like NOD32 does) or does it stop operating completely?

Hi, when Norton products expire they completely stop working. You will not be able to update or even run scans. If you will not be renewing and are going to be using another product be sure to uninstall NAV from Windows Control Panel. Then reboot

and download and run the Norton Removal tool before you install your new product.

Task A7 - English ST:

Norton toolbar extension in Google chrome

Hello, I am having an issue with the NORTON extension toolbar slowing and freezing webpages when loading. I updated to 360you about 3 months ago and have not had a problem until about 7-10that days ago. Page freezes when loading, I can't scroll or interact with page at all. If I open a new tab or go to different tab already open and come back, original page is fully loaded and operates fine. I even bought a new router because, I bought a new laptop and problem was more prevalent. Thought maybe the wireless connection was acting up. I upgraded from a duo wireless N router WNDR3300 to an R6300 Widow Router 802.11ac Dual Band Gigabit. I still had issue with both PCs: desktop with Ethernet cable router connection, and NEW laptop with wireless connection. Both PCs are running Windows 7 with NORTON 360. I disabled NORTON extension in chrome settings and problem goes away. All my products are up to date. I leave PC on all the time with automatic updates enabled. I would like to know if there is a fix for this, or maybe NORTON doesn't know about it yet. NORTON 360 version: 6.3.0.14 Desktop: win 7, 4G of memory, 500Gb hard drive, AMD Agglomerate II X2 255 processor 3.10 GHz. Laptop: win 7, 8G of memory, 750Gb hard drive, 2640M(17) processor. Thanks for help.

Try this and see if it fixes the freeze/hang issue, until a more permanent fix comes along: 1) Open Chrome and type "about:plugins" in the address bar. 2) When the plugins page opens, click Details in the upper right side of the page. 3) Near the top of the page will be 2 files for Adobe Flash Player. 4) Find the file with the location that ends in Pepper Flash or pepflashplayer.dll. 5) Click Disable for this file and restart the Chrome browser. 6) Go to: get. adobe .com/ flashplayer / and download the latest version of Flash Player. 7) Once it installs, restart Chrome. Try a few websites where Chrome would freeze and see if the problem is now solved.

Task A8 - English ST:

trojan in a computer

Good Evening, I need some help. A friend of mine has trojan viruses on her computer. She has no anti-virus protection on it. If she were to purchase Norton will it remove these trojan viruses from her computer? Thank you for your help.

In that case I would suggest registering at one of the malware removal sites. Follow their instructions and you should get cleaned up.

`http://www.bleepingcomputer.com/`   `http://www.geekstogo.com/forum/`
`http://www.cybertechhelp.com/forums/` `http://support.emsisoft.com/forum/`
`6-help-my-pc-is-infected/ http://forums.whatthetech.com/`

Task A9 - English ST:

Activation key for NMS bought through android market

I bought NMS through the market in the end of Jan. Recently, my ROM corrupted and bricked my phone. I managed to recover, however, NMS is missing from my owned apps on market. I can't find the activation code to activate via the norton.com/nms link. Another minor annoyance is that my main Norton account is registered to my email address, though NMS doesn't appear on my Norton account at all. The only record I have of my purchase is through the purchase history in my android market account and a copy of the invoice. Where should I go to get my activation key or do I need or hassle Google to get my apps list fixed

Hi. Just to let you know when you purchase the upgrade to NMS from Android Market it will not show in your Norton Account product list. There is not a activation product key for the Market version of NMS. If you purchase a retail version of NMS it too will not register as a Norton product in your Norton Account. This happens as there is no link to it when activated with the product key included with pack. It's different when you purchase Norton Mobile Security from the Norton Online Store. The purchase will register as a product in your Norton Account with the product key used to activate it which will also be included. And the link www.norton.com/nms

link is used to download the apk file. It's also used to install NMS for the retail and Norton Online Store versions of Norton Mobile Security. I hope that info helps. Cheers.

Task A10 - English ST:

Norton Online R6025 Runtime error

Hello, now I installed Norton Online Family. On 3 computers have I no problem. But on one is generated Runtime error after starting every internet browser (IE, Chrome, Firefox, Opera). Also starting of every browser is very slow. Exact message text is: Runtime Error! Program:
C:\\Program Files\\Norton Online\\Engine\\2.3.0.7\\ccSvcHst.exe R6025 -pure virtual function call. Operating System: Windows XP SP3 Installed Norton products: Norton Safety Minder 2.3.0.17, Norton Online Framework 2.3.0.7, Norton Online Datastore 2.3.0.7 Text in Applications Event logger : The description for Event ID (35) in Source (NOF). The local computer may not have the necessary registry information or DLL files required reports to display messages from a remote computer. To obtain this description, you can use the / AUXSOURCE = flag . For more information, see Help. To event is attached the following information: NOF. Thanks for help.

Hi , thanks for the update. Please check the link below where shows the security products that the program is not compatible with. `http: //www.symantec.com/norton/support/kb/web_view.jsp?wv_type=public_web& ssfromlink=true&docurl=20090508145137EN` Thanks.

Task A11 - English ST:

Purge in progress for 2-1/2 days... help!

Hi. I need someone on the Symantec side to get the Purge process to finish up on one of my 2 computers associated with NOBU. It's the third day now that the "purge in progress" message is showing. Logout and reboot have no effect. This backup profile is for my primary business computer and it's my offsite backup. No backups have run since the purge, and if I need access to a backed-up file, I am out of luck. Reinstallation is not an option. I won't have the time for a week to go through that. Thanks much.

Hi that seems to have done the trick. I had turned off the service entirely for a week, and then thought I'd try it again today – and it's working. Then I checked the message board, and found your note. So we can call this SOLVED. Thanks

Task A12 - English ST:

Norton Utilities 15.0.0.124 Problems With "Clean Your Registry"

Whenever I run this with a deep scan it screws up my CanoScan Toolbox 5.0 application for my CanoScan 4400F. So when I try to scan a document it cannot find the scanner. The only solution is for me to install the program and reinstall it again. This also happened under NU 14.

I finally isolated the problem and found that NU was deleting items pertaining to the Toolbox both in appdata and in the registry. I added this to the exclusion list. However, NU should not have been deleting this. Perhaps it is a bit too aggressive.

Task B1 - English ST:

IS this a virus

An Icon shows on my computer and it is rminstall[1].exe. I cannot find a file for this on drive c. Is this a virus and if so how do I get rid of it. I have performed a full system scan and only 4 cookies were detected.

Hello! Do you happen to have Registry Mechanic installed on your system? If you do, it belongs to that program.

Task B2 - English ST:

Can Norton Antivirus License Be Transferred From One Computer To Another?

Hi, I have a Norton Antivirus license that extends up to February 2013. However I had to change my laptop recently and my Norton Antivirus license was purchased for my old laptop. I don't use my old laptop any more. Is there any way I can transfer the same license to my new laptop so that I can leverage the full benefit of my subscription? Any information about this would be deeply appreciated. Thanks and Regards.

Hi, the best option would be to contact Customer Support and ask them. They are very good at this type of thing. Live Chat is usually the quickest. You can reach them at `http://www.norton.com/chat`. The other option is to sign in to Norton Management with your Norton Account details and click on the My Account tab. Click the trash can icon, Then you will be able to install on your new laptop. Remember that if it comes with any other security program installed it will need to be thoroughly removed using that programs removal tool.

Task B3 - English ST: I want to stop Norton 360 from blocking images in emails on Windows Live Mail on Windows 7

Just about every email with images in the html code are blocks and I have to manually OK each time to see them. This is terribly annoying. I have tried turning off anti-spam and email scanning but it's still blocking pictures. How do you permanently disable this 'feature' from Norton 360.

Windows Live Mail, like most current email clients, blocks images by default. This is primarily to protect you from spam. Viewing an image in a spam message will allow the spammer to know that your email address is ̈liveänd ready for heavy bombardment. See the following Microsoft article: View images that Windows Mail blocked

Task B4 - English ST:

IE9 and NIS toolbar question

I've noticed that if I open IE9, the Norton Toolbar and Identity Safe Icon show up. If I pin a website to the desktop (Like this Community Forum), the Norton Toolbar and Identity Safe Icons do not show up??? By using the pinning feature, am I bypassing NIS security? Have I missed something? Is there a way I can get my username and password to be entered from Identity Safe? I'm using Windows 7 and NIS 2102 using IE9 (I haven't tried it with Firefox. I just upgraded to version 13 and I am still setting up options.

Hi, what you are seeing is correct. Opening a site that is pinned to the task bar opens an isolated IE page and the Norton Toolbar does not display. I had this site pinned for a short period until I realised the ID Safe wouldn't fill to sign in. Yes, you are still protected. So now I have limited sites pinned and instead rely on bookmarks (IE calls them favorites). As far as I am aware, you can't pin sites with Firefox or Chrome.

Task B5 - English ST:

Unable to get NIS 2012 disk to automatically start

I need to install via disk so I need to see if anyone can help me with this. I logged in as admin account I placed the disk in in the CD drive. I hear "running" sounds, but nothing pops open. When I go to my computer, I can see NIS in the CD Drive. If I right click on it there, it will open up and give me the menu to start the installation. My concern is why it doesn't automatically pop up the installation menu when the disk is entered in the disk drive? Could it have to do with the auto run feature off?

Yes, it's definitely the autorun feature. For security reason, it's best for it to be off. However, as you notice, CD and DVD disk setup files no longer start by themselves.

So just do the manual start of the installation, instead of messing with turning autorun on and off. It will work 100% the same.

Task B6 - English ST:

NIS2012 and Adguard

Thought I would try this program at http://adguard.com/en/welcome.html but when I download adguardinstall.exe NIS immediately deletes it saying that it is a threat. I wonder why it would be a threat and wonder if anyone here uses it?" Many thanks.

Hi, reading about the program on their site it appears to have an active scanner. Real-time scanners must work alone. Norton also has a real time scanner. If you were able to download and install the second program the minimum result would be diminished protection. This is because the two fought over what to do with whatever they found that was suspect. The worst possible case would be that they effectively cancel each other and you would have no protection. I cannot recommend that you consider using this program while you have Norton installed. Stay well and surf safe

Task B7 - English ST:

Norton Ghost ALWAYS says "At Risk"

I have used Ghost forever and it has always worked great and been simple. Now I am using 15.0.1.36526 on Windows 7 32 bit SP1. The backups go to an external WD RAID unit, I don't recall the model but it has twin 2TB drives in a mirrored array. I have a single drive (C:) in my machine and a single backup job that always runs at lunch. No problems there. But Ghost ALWAYS says "At Risk" (systray icon and Ghost Home screen) so I always have to open Ghost and manually check... Was there a problem with today's backup??? I occasionally have an external drive that I plug in from time to time. On the status screen I see a table with the C: drive and the BU jobs by day. I also see the WD BU unit listed as [My Book 2x2TB] (F:) listed with an X in a red hexagon. On the advanced screen I see my C: drive with a green circle and check 2 other C: drives with questions marks. I also see System Reserved (D:\\) with question mark and(E:\\) with question mark. Lastly, I see the [My Book 2x2TB]

(F:) listed with an X in a red hexagon. All I care about is the first C: drive in this list. What do I need to do to make the "At Risk" condition go away without risking deleting my backups ?

Open Ghost and click the big button "Advanced". Right click on the other drives and select "customize status reporting". Then select "no status reporting". Leave the C drive and the destination drive the way they are. You want reporting on the C drive and the destination drive should already be turned off and listed as "backup destination".

Task B8 - English ST:

Full System Scan error

I recently installed Norton Internet Security 2012 on my PC running Windows 7. When I open Norton and select the option to do a full system scan, the scan window shows (in red): Error During Scan. This happens after rebooting my system. If I again select the full system scan option without rebooting, nothing happens (I don't get the scan window or the error message). If I select Quick Scan, that works ok. I previously had Norton Internet Security 2011 installed and I did not have any problems with the full system scan. Thanks for your help.

Hi, Welcome to Norton Community! Sorry for the trouble you have with full system scan. Please check in Security History and verify if there are any errors logged for the Full System Scan. You may be able to find the error number and module from the details, which might be helpful for us in determining the issue. In meantime, you can restart the computer in Safe Mode, & double-click Norton icon to run a full system scan. Check if it runs the scans or detects any threats. If the problem persists, restart the computer to normal mode and try Norton Power Eraser: `http://www.norton.com/npe`. If the NPE also didn't detect any problems, then you may have to uninstall/reinstall your NIS 2012 using Norton Remove & Reinstall tool: `http://www.norton.com/nrnr`. Let us know how it goes. Thanks.

Task B9 - English ST:

Ghost 15 restore failure

Situation: strange behaviour of windows 7 prof 64b (programs wouldn't start, no network access) after trying to unlock my mobile phone. This led me to attempting to restore drive C: Restore process by boot-CD starts fine from external USB-drive but after 2/3rds of completion it stops and causes this message: "error EA390013: At end of something" I tried different backup sets (backup was every 2 hours), with and without deletion of the target drive. Computer is a Sony Vaio notebook with double SSD drive in RAID 0 configuration. I really hope, someone can give me some good advice. I need my computer urgently.

While booted to the restore disk, wipe the disk you are restoring clean. Open the command prompt and follow the instructions here, starting with step # 3. When you get to # 11, don't worry about creating a new partition as the Ghost restore will do this. `http://www.sevenforums.com/tutorials/ 52129-disk-clean-clean-all-diskpart-command.html`. ENSURE THAT YOU ARE WIPING THE CORRECT DISK AND THAT YOU DO NOT WIPE THE DISK WITH THE IMAGE FILE ON IT!!! Once done, reboot the computer back into the SRD, and attempt the restore.

Task B10 - English ST:

USB Flash Drive recovery

Hello, I run my N360 (ver. 20.2.0.19) backup to a Sandisk Cruzer USB Flash Drive (16GB). There are also other files (mostly photos) on the USB flash drive that are not included in the N360 backup. When I put the flash drive in the computer yesterday to perform a backup the flash drive came up as empty (no files). I do not own Norton Ghost. My 2 questions are: How can I go about recovering the files from my flash drive (assuming they are still there)? Can 360 or Ghost help in doing so? Thank you.

Welcome, how they can be recovered will depend on how they went missing. Has any physical or electrical trauma occurred? Have you installed the flash drive and

rebooted? Thanks.

Task B11 - English ST: How best to install a new copy of NIS 2012 over an existing but time limited installation?

Hi. I had Norton Internet Security 2011 installed on my computer. I took advantage of the free upgrade to NIS 2012. This runs out in a few days' time. I have a brand new retail copy (boxed) of NIS 2012. What is the best way ti install the new license? Should I un-install the current NIS 2012 then install the new copy? I would prefer to keep my existing settings. So can I re-install the new copy over the existing version? Or thirdly is there a way I can simply input the new product key and keep the existing installation with an extended time period? Thanks for any help and advice.

Hello, you do not need to uninstall/reinstall. All you need to do is enter your new Activation key in to your current subscription before it expires. Just wait until your subscription is a day or two before it expires, then open NIS and click on Renew.

Task B12 - English ST:

I have uninstalled it properly but the license has not been deactivated

I have used the Norton Removal tool thing to remove Norton 360 from my old HDD then installed it a new one But I have come to find that I used to have 3 licenses available but now I only have 2. I want to have the other one back. Also the Live chat support is of absolutely no use, because it won't even load up the chat page.

Go here and sign in with your Norton account `https://manage.norton.com/Account`. Click the large tab "My Account" Then under your license you will see the activations used and a "trash can" icon you can click to remove the license from that system.

Appendix E

MT output - Group A and B

Task A1 - German MT output:

W32.Downadup.B

Hallo, My Mac infiziert wurde mit W32.Downadup.B. Jede Ressource sagt, dass diese Infektion ist nur für PCs und bietet Lösungen auf PCs, um es zu lösen. Ich benötige, um es zu entfernen, denn es ist mein Browser verlangsamen und anderen Computern im Netzwerk. Ich möchte nur ohne lächerlich Ladezeiten netflix ansehen.

Ich denke, auf einem Versuch Sie erfolgreich entfernt die Datei (mit dem Befehl nach Erfolg würde zeigen rm "Datei nicht gefunden" Hoffentlich wird dies verhindern, dass die wiederholten Berichte, aber wenn nicht, lassen Sie mich wissen.

Task A2 - German MT output:

Kopieren, clip Board für Red erkannte Bedrohungen Bildschirm"

In der Vergangenheit, wenn ich mit Norton NIS 2012 zu clip kopiert die Tracking-Cookies wäre aufgelistet, wenn ich sie zu einem Notepad eingefügt. Jetzt mit der neuen Version nach Re- Abonnements, die nicht sagen NIS 2013 nur sagt (Norton Internet Security) ver. 20.2.1.22. Die Tracking-Cookie Informationen aus den einfügen. Ich verwende diesen Informationen blockiert alles, was in meinem IE9 Browser-Datenschutz Websites. Ist das normale jetzt mit der ver. 20.2.1.22? Ich habe versucht, mit dem

Startmenü deinstallieren und nach, dass die Verwendung des Norton _ Removal _ Tool. Nach der Neuinstallation alles funktioniert die gleichen, keine spezifischen Tracking-Cookie Informationen über die einfügen.

Nein, es ist nicht normal; es ist ein Fehler in der neuen Norton 2013 Produkte. Der in der Dropdown-Liste "Anzeigen" alle "Dateiinfo" Windows wurde neugestaltet der Windows 8 Touch zu unterstützen. Eine unbeabsichtigte Folge dieser Konzeption ändern, die "in Zwischenablage kopieren" in diesem Dialogfeld nicht mehr Funktionen als erwartet. Um das Problem zu arbeiten, müssen Sie einen Quick Scan ausführen. Wählen Sie den Link "Exportergebnisse", um die Prüfung unten angezeigten Ergebnisse in eine Textdatei: Öffnen Sie diese neu in der Textdatei gespeichert. Sie sollten nun sehen, die detaillierte Tracking-Cookie Informationen. Dies hilft Ihnen mit Ihrer Entscheidung, ob sollten Sie jedes der aufgeführten Websites blockieren. Hoffen, dass dies hilft.

Task A3 - German MT output:

File Guard nicht geöffnete Dateien sogar mit den richtigen Administratorkennwort.

Der Titel ziemlich viel sagt alles. Ich habe, die Datei zu schützen, meine geschützten Dateien nicht öffnen, weil er sich selbst mit der richtige Name und Kennwort des Administrators. Dies geschieht vor ein paar Tagen vor, und es gut funktioniert. Alle Ideen?

Hallo, das möglicherweise ein Problem mit File Guard Befüllen der Benutzername Bereich mit den vollständigen Namen anstelle der kurzen (aka, POSIX) Benutzernamen. Um festzustellen, die kurze Benutzername: 1. Öffnen Sie / Anwendungen / Utilities / Terminal.app 2. Geben Sie: Whoami 3. Drücken Sie zurückkehren. Ihrer kurzen (POSIX) Benutzername wird durch die whoami Befehl angezeigt werden. Verwenden Sie diesen Namen, wenn File Guard Anweisungen für Ihr Administratorkennwort. Bitte lassen Sie uns wissen, ob dies hilft. Vielen Dank!

Task A4 - German MT output:

Besorgt über versteckte Virus auf NIS 20.3.0.36

Vor zwei Tagen geklickt haben, als ich auf einen Link in einer E-Mail dachte ich war von einem Freund. Sobald es eine Seite geöffnet, ich habe bemerkt, dass die E-Mail nicht sent by a friend... Sie hatte, dass ein Hacker. Ich habe drei Prüfungen mit NIS und eine vollständige Systemprüfung mit freien Malwarebytes Anti-Malware. Alle sind mit sauberen Ergebnisse, und meine 64-Bit Windows 7 Computer ist normal. Gibt es etwas weiter ich tun sollten, um sicherzustellen, dass ein Hacker nicht zurückgegangen ist eine Datei auf meinem System? Ich bin zögerlich zur Verwendung der Norton Power Eraser Tool. Mein Dank...

Hallo, Einträge in Norton Produktmanipulationsschutz Datensatz jeden Versuch einer externen Agent für den Zugriff auf einen Norton Datei oder Prozess. In fast allen Fällen verursacht werden, die Ereignisse von legitimen Programmen oder Windows Prozesse, und nichts zu befürchten. Das stimmt der Einträge, die Sie veröffentlicht haben. Das sind alles legitime Dinge, die auf Ihrem Computer ausgeführt wird. Norton ist einfach dafür zu sorgen, dass nichts, nicht einmal Windows, Probleme verursachen können durch die Norton Vorgang. Es gibt keine Hinweise darauf, was Sie etwas bösartigen in veröffentlicht haben.

Task A5 - German MT output:

Benötigen SIE HILFE SCHNELLE!!

Wenn ich ctrl ALT löschen und klicken Sie auf Start der Aufgabe Manager und fahren Sie unter Prozesse. Was ist Norton heißen? Ich benötige beides so schnell wie möglich bitte!!!

Hallo, sind Sie sicher, dass Norton blockiert TeamViewer? Öffnen Sie die Programmsteuerung Fenster in Norton ("Einstellungen" > "Netzwerk" > "Intelligente Firewall" > "Programmsteuerung). Wenn Sie TeamViewer aufgelistet wie "Auto", dann Norton ist nicht blockieren, und das Problem liegt anderswo. Sie können jedoch, entfernen Sie die TeamViewer Eintrag aus der Liste, um mögliche Fehlkonfiguration automatisch reparieren. Stellen Sie zuerst sicher TeamViewer wird nicht ausgeführt. Klicken Sie dann auf die TeamViewer Eintrag und klicken Sie auf "Entfernen", "Übernehmen" und "OK". Beenden von Norton und starten Sie TeamViewer. Norton wird automatisch neue Regeln für das Programm konfigurieren, und es sollte. Sie können die Norton

Prozess töten. Es ist immun gegen außerhalb -manipulation.

Task A6 - German MT output:

Was passiert, wenn eine NAV Lizenz abläuft?

Ist es nur konnte nicht aktualisiert werden (wie NOD32), oder ist es völlig aufhören?

Hallo, wenn Norton Produkte laufen sie völlig aufhören zu funktionieren. Sie werden nicht in der Lage sein, zu aktualisieren oder sogar Scans ausführen. Wenn Sie nicht die Erneuerung und werden mit einem anderen Produkt sicher sein, deinstallieren Sie NAV über die Systemsteuerung in Windows. Starten Sie dann und Download und Ausführung des Norton Removal Tool installieren, bevor Sie Ihr neues Produkt.

Task A7 - German MT output:

Norton-Symbolleiste Erweiterung in Google Chrom

Hallo, ich habe ein Problem mit der Norton Erweiterung Symbolleiste beim Laden von Webseiten und das Einfrieren zu verlangsamen. Ich aktualisiert, 360you über 3 Monaten und nicht ein Problem bis über 7-10that Tagen. Beim Laden der Seite Abstürze blättern, kann ich nicht auf alle oder Interaktionen mit Seite. Wenn ich eine neue "oder anderen Registerkarte bereits geöffnet und wieder, den ursprünglichen Seite ist vollständig geladen und funktioniert gut. Ich selbst eine neue Router gekauft haben, weil ich eine neue Laptops und Problem weiter verbreitet wurde. Dachten vielleicht war die WLAN-Verbindung. Ich ein Upgrade von einer Duo drahtlosen N Router WNDR3300 zu einem R6300 Widow Router 802.11ac Dual Band Gigabit. Ich hätte noch Problem mit sowohl PCs: Desktop mit Ethernet-Kabel Router Verbindung und NEU Laptop mit drahtlosen Verbindung Sowohl PCs mit Windows 7 mit Norton 360. Ich deaktiviert Norton Erweiterung im Chrome-Bericht Einstellungen und Problem geht. Alle meine Produkte auf dem neuesten Stand sind. Ich überlasse PC auf alle Zeit mit automatischen Updates aktiviert. Ich möchte wissen, ob es eine Lösung für dieses, oder vielleicht Norton ist es noch nicht kennen. Norton 360 Version: 6.3.0.14 Desktop: Win 7, 4 GB Arbeitsspeicher 500Gb Festplatte, AMD Agglomerate II X2

255 Prozessor 3.10 GHz. Laptop: Win 7, 8G Arbeitsspeicher 750GB Festplatte 2640M (17) Prozessor. Vielen Dank für helfen.

Versuchen Sie diese und zu sehen, ob es behebt das Einfrieren / hängen bleiben, bis eine dauerhaftere Lösung kommt:

1) Öffnen Sie Chrome und geben Sie "über: Plug-ins" in die Adressleiste ein.

2) Wenn die Plug-ins Seite geöffnet ist, klicken Sie auf "Details" in der oberen rechten Seite der Seite.

3) Nahen ganz oben auf der Seite 2 Dateien für Adobe Flash Player.

4) Suchen Sie die Datei mit dem Standort, endet in Pepper Flash oder pepflash-player.dll.

5) Klicken Sie auf "Deaktivieren" für diese Datei und starten Sie den Chrome Browser.

6) Go to: Get. adobe .com / Flashplayer /, und laden Sie die neueste Version von Flash Player.

7) Wenn es installiert, starten Sie Chrome.

Versuchen Sie es einige Websites, auf denen Chrome eingefroren und sehen, ob das Problem gelöst ist.

Task A8 - German MT output:

Trojaner in einen Computer

Guten Abend, ich muss einige helfen. Ein Freund von mir hat Trojaner Viren auf ihrem Computer. Sie hat keine Virenschutz. Wenn sie zum Erwerb von Norton wird es entfernen Sie diesen Trojaner Viren aus ihrem Computer? Vielen Dank für Ihre Hilfe.

In diesem Fall schlage ich vor der Registrierung auf einem der Malware entfernen. Http: / / www.bleepingcomputer.com / http: / / www.geekstogo.com / Forum / http: / / www.cybertechhelp.com / forums / http: / / support.emsisoft.com / Forum / 6-help-my-pc-is-infected / http: / / forums.whatthetech.com /. Folgen Sie den Anweisungen und sie sollten sich bereinigt.

Task A9 - German MT output:

Aktivierungsschlüssel für NMS durch Androiden Markt gekauft

Ich gekauft habe NMS durch den Markt in das Ende von Jan. Vor kurzem hat mein ROM beschädigt und bricked meine Telefon. Ich geschafft hat, sich zu erholen, aber NMS fehlt in meinem Besitz apps am Markt. Ich kann nicht den Aktivierungscode Aktivierung über das norton.com / nms Link. Ein weiteres kleines Ärgernis ist, dass mein Norton Account registriert ist, meine E-Mail-Adresse, obwohl NMS doesn 't appear in meinem Norton Account. Nur Protokoll, ich habe mein Produkt ist durch den Kauf Geschichte in meinem Androiden Markt und eine Kopie der Rechnung. Wo sollte ich meinen Aktivierungsschlüssel oder benötige ich oder Aufwand Google meine apps Liste behoben?

Hallo. Nur, damit Sie wissen, wann Sie das Upgrade auf NMS von Android Market wird es nicht in Ihrem Norton Account Produkt anzeigen. Es ist nicht eine Aktivierung Produktschlüssel für den Binnenmarkt Version von NMS. Wenn Sie eine Einzelhandelsversion von NMS es auch nicht registrieren Sie sich als Norton Produkt in Ihrem Norton Account. Dies geschieht, denn es ist aktiviert, wenn es keine Verbindung mit dem Produktschlüssel mit Pack enthalten. Ihre verschiedenen beim Kauf von Norton Mobile Security vom Norton Online Store. Der Kauf registrieren wird in Ihrem Norton Account ein Produkt mit dem Produktschlüssel zu aktivieren, die verwendet werden ebenfalls enthalten. Und den Link www.norton.com / nms Link APK wird verwendet, um die Datei herunterzuladen. Die für die Installation NMS für den Einzelhandel und Norton Online Store Versionen von Norton Mobile Security. Ich hoffe, dass Informationen hilft. Beifall.

Task A10 - German MT output:

Norton Online R6025 Laufzeitfehler

Hallo, ich jetzt Installation von Norton Online Family. Auf 3 Computern habe ich kein Problem. Aber auf einem wird generiert, Runtime Fehler nach dem Starten jeder Internet-Browser (Internet Explorer, Chrome, Firefox, Opera) Auch alle Browser ist sehr langsam. Exakte Nachricht Text ist: Runtime Error! Programm: C:\\Program

Files\\Norton Online\\Engine\\2.3.0.7\\ccSvcHst.exe R6025 -pure virtueller Funktionsaufruf. Betriebssystem: Windows XP SP3 installiert Norton Produkte: Norton Safety Minder 2.3.0.17, Norton Online Framework 2.3.0.7, Norton Online Datastore 2.3.0.7 Text in Anwendungen Event Logger: Die Beschreibung für Event ID (35) in Quelle (Nof). Der lokale Computer hat unter Umständen nicht die erforderlichen Registrierungsinformationen oder DLL Dateien geforderten Berichte anzeigen, Meldungen von einem Remote-Computer." Diese Beschreibung verwenden, können Sie die / AUX-SOURCE = Flagge. Weitere Informationen finden Sie in der Hilfe. Ereignis verbunden ist, die folgenden Informationen: Nof. Vielen Dank für helfen.

Hallo, vielen Dank für die Aktualisierung. Überprüfen Sie den Link unten, wo zeigt die Sicherheitslösungen, die das Programm ist nicht kompatibel with. `http://www.symantec.com/de/de/Norton/support/kb/web_view.jsp?WV_type=` `public_web&ssfromlink=true&docurl=20090508145137EN` Vielen Dank.

Task A11 - German MT output:

Bereinigung in Fortschritt für 2-1 / 2 Tage... zu helfen!

Hallo. Ich benötige jemand auf der Symantec Seite, um den Löschvorgang Prozess abgeschlossen ist, bis auf eine meiner 2 Computer im Zusammenhang mit Nobu Der dritte Tag jetzt bereinigen, die "Fortschritte" Meldung angezeigt ist. Abmelden und neu gestartet haben keine Auswirkungen. Dieses Backup Profil ist für mein Unternehmen Computer und es ist meine Offsite-Backups. Sind keine Backups ausgeführt werden, da die Bereinigung, und wenn ich benötige Zugang zu einer gesicherten Datei, ich bin Pech. Neuinstallation ist keine Option. Ich habe nicht die Zeit für eine Woche durch. Vielen Dank.

Hallo, scheint das Kunststück getan haben. Ich hatte deaktiviert den Dienst völlig für eine Woche, würde ich dachte, und versuchen Sie es erneut heute – und es funktioniert. Dann habe ich die Meldung, und Ihr Hinweis. So können wir nennen diese SOLVED. Vielen Dank

Task A12 - German MT output:

Norton Utilities 15.0.0.124 Probleme mit "Clean Your Registry"

Wenn ich dies mit einem tiefen Scan ausführen, das Schrauben meine CanoScan Tool-box 5.0 Anwendung für meine CanoScan 4400F. Wenn ich also versuchen, ein Dokument, mit dem sie die Scanner nicht finden können. Die einzige Lösung ist für mich, installieren Sie das Programm und installieren Sie es erneut. Dies geschah auch unter NU 14.

Abschließend möchte ich das Problem isoliert und festgestellt, dass die NU war Löschen von Elementen in Bezug auf die Toolbox sowohl in AppData und in der Registrierung. Ich diese in die Ausschlussliste hinzugefügt. Aber NU sollten nicht gelöscht wurden. Vielleicht ist es ein bisschen zu aggressiv.

Task B1 - German MT output:

Ist dies ein Virus

Zeigt ein Symbol auf meinem Computer, und es ist rminstall [1] .exe Ich eine Datei nicht finden kann für diese auf Laufwerk c. Ist dies ein Virus und wenn ja, wie kann ich es loszuwerden. Ich habe eine vollständige Systemprüfung durchgeführt und nur 4 Cookies erkannt wurden.

Hallo! Haben Sie Registry Mechanic auf Ihrem System installiert? Wenn Sie das tun, es gehört zu diesem Programm.

Task B2 - German MT output:

Kann Norton Antivirus License Be übertragen von One Computer So Anderer?

Hallo, ich habe ein Norton Antivirus Lizenz, die bis Februar 2013. Allerdings hatte ich meinen Laptop kürzlich zu ändern, und mein Norton Antivirus Lizenz erworben wurde für meinen alten Laptop. Ich verwende meiner alten Laptop mehr. Gibt es eine Möglichkeit, mit der gleichen Lizenz kann ich meinen neuen Laptop nutzen können, so dass ich mein Abonnement den vollen Nutzen? Informationen über das wäre sehr geschätzt. Dank und mit freundlichen.

Hallo, die beste Option wäre, wenden Sie sich an den Kundenservice und sie bitten. Sie sind sehr gut auf diese Art der Sache. Live-Chat ist normalerweise die schnellste. Sie erreichen können Sie unter http: / / www.norton.com / Chat. Die andere Option ist, um sich bei Norton Management mit Ihrem Norton Account Details und klicken Sie auf die Registerkarte "Mein Account". Klicken Sie auf das Symbol "Papierkorb kann, dann können Sie auf Ihrem Laptop zu installieren. Denken Sie daran, dass, wenn es darum geht, die mit jedem anderen Sicherheitsprogramm installiert werden müssen, gründlich mit, dass Programme Entfernungsprogramm entfernt.

Task B3 - German MT output:

Ich möchte Norton 360 von der Blockierung Bilder in E-Mails auf Windows Live Mail auf Windows 7

Nur über alle E-Mails mit Images in der HTML-Code angezeigt werden blockiert, und ich muss manuell "OK" jedes Mal, sie zu sehen. Dies ist besonders ärgerlich. Ich habe versucht deaktivieren Antispam- und E-Mail-Prüfung, aber es ist noch blockiert Bilder. Wie können Sie diese Funktion deaktivieren "von Norton 360.

Windows Live Mail, wie die meisten aktuellen E-Mail-Clients, Images standardmäßig blockiert. Dies ist in erster Linie zu schützen Sie vor Spam. Anzeigen eines Image in einer Spam-Nachricht wird es ermöglichen, die Spammer zu wissen, dass Ihre E-Mail-Adresse ist "lebenden" und bereit für schwere Bombardierung. Beachten Sie die folgenden Microsoft Artikel: Bilder, die Windows Mail blockiert

Task B4 - German MT output:

IE9 und NIS Symbolleiste Frage

Ich habe bemerkt, dass, wenn ich IE9 öffnen, wird die Norton-Symbolleiste und Identity Safe Icon zeigen. Wenn ich eine Website setzen auf den Desktop (wie dieser Community Forum), die Norton Toolbar und Identity Safe Icons nicht angezeigt werden??? Mithilfe der Funktion, bin ich übrigens unter Umgehung der NIS Sicherheit? Ich habe etwas verpasst? Gibt es eine Möglichkeit erhalte ich mein Benutzername und Kennwort eingegeben werden von Identity Safe? Ich bin mit Windows 7 und NIS 2102 mit IE9

(ich noch nicht versucht es mit Firefox. Ich nur ein Upgrade auf Version 13 und ich bin immer noch die Einrichtung Optionen.

Hallo, was Sie sehen, ist richtig. Eröffnung einer Site befindet, die an der Taskleiste öffnet eine isolierte IE Seite und wird die Norton-Symbolleiste nicht angezeigt. Ich hatte diese Website für einen kurzen Zeitraum ruhen, bis ich die ID Safe wouldn 't ausfüllen zu unterzeichnen. Ja, Sie sind immer noch geschützt. So habe ich jetzt beschränkt sich auf Websites und stattdessen auf Lesezeichen (IE nennt sie Favoriten). Soweit ich weiß, können Sie nicht die PIN Websites mit Firefox oder Chrome.

Task B5 - German MT output:

Datenträger konnte nicht abgerufen werden automatisch zu starten, um NIS 2012

Ich also über Datenträger installieren müssen, um zu sehen, ob jemand muss ich mich mit diesem helfen können. Ich angemeldet als Admin-Konto ich die Datenträger in das CD-Laufwerk ein. Ich höre "Ausführen" klingt, aber nichts wird geöffnet. Wenn ich meinen Computer, kann ich sehen, NIS im CD-Laufwerk befinden. Wenn ich richtig, klicken Sie auf, es wird geöffnet und geben Sie mir die ", um die Installation zu starten. Meine Sorge ist, warum es nicht automatisch das Installationsmenü angezeigt, wenn der Datenträger in das Laufwerk? Könnte es zu tun haben, mit der automatisch ausgeführt Funktion ausschalten?

Ja, es ist definitiv die selbstinstallierende Funktion. Für Sicherheit Grund, ist es am besten, dass es deaktiviert sein. Aber, wie Sie feststellen, CD und DVD Datenträger Setup Dateien selbst nicht mehr starten. So führen Sie die manuellen Start der Installation, anstatt auf selbstinstallierende Handel mit aktivieren und deaktivieren. Es funktioniert 100% der gleichen.

Task B6 - German MT output:

Adguard NIS2012 und

Ich möchte versuchen, das Programm unter http: / / adguard.com / en / welcome.html adguardinstall.exe NIS herunterzuladen, aber wenn ich sofort löscht es sagen, dass sie

eine Bedrohung. Ich frage mich, warum es wäre eine Bedrohung und frage mich, wenn jemand hier verwendet? Vielen Dank.

Hallo, Lesung über das Programm auf ihrer Website zu haben scheint es eine aktive Scanner. Allein in Echtzeit Scanner arbeiten müssen. Norton hat auch eine wirkliche Zeit Scanner. Wenn Sie herunterladen und installieren konnten die zweite Programm für das Ergebnis verringert würde, Schutz. Dies ist, weil die beiden gekämpft, was zu tun ist, mit dem, was sie gefunden, die verdächtig war. Der schlimmste Fall wäre, dass sie tatsächlich abbrechen und Sie haben keinen Schutz. Ich kann nicht empfehlen, dass Sie dieses Programm verwenden, während Sie Norton installiert. Bleiben Sie gut und sicher im Internet surfen

Task B7 - German MT output:

Norton Ghost IMMER sagt "Gefährdet"

Ich habe Ghost verwendet, und es hat immer für große und einfach. Jetzt bin ich mit 15.0.1.36526 auf Windows 7 32-Bit SP1. Die Backups auf einer externen WD RAID Einheit, ich kann mich nicht erinnern, das Modell, aber es hat zwei 2 TB Laufwerke in einer gespiegelten Array. Ich habe ein einzelnes Laufwerk (C:) in meinem Computer und einer einzigen Backup-Auftrag, der ständig beim Mittagessen. Keine Probleme. Aber Ghost IMMER sagt "Gefährdet" (in der Systemleiste und Ghost Home Bildschirm), damit ich immer öffnen, um Ghost und manuell aktivieren... War es ein Problem mit todays Backup??? Ich gelegentlich ein externes Laufwerk anschließen, dass ich von Zeit zu Zeit. Über den Status Bildschirm sehe ich eine Tabelle mit dem Laufwerk C: Und BU Jobs nach Tag. Ich sehe auch die WD BU Einheit als [Mein Book 2x2TB] (F:) mit einem X aufgeführt, die in einem roten hexagon. Über die erweiterten Bildschirm sehe ich mein Laufwerk C: Mit einem grünen Kreis und aktivieren Sie 2 anderen C: Laufwerke mit Fragen. Ich sehe auch System vorbehalten (D:\\) mit Fragezeichen und (E:\\) mit Fragezeichen. Schließlich sehe ich die [Mein Book 2x2TB] (F:) mit einem X aufgeführt, die in einem roten hexagon. Ich interessiert, ist das erste Laufwerk C: In dieser Liste enthalten. Was muss ich tun, um die "Gefährdet" Bedingung verschwinden, ohne zu riskieren, meine Backups löschen?

Öffnen Sie den Ghost und klicken Sie auf die Schaltfläche "Erweitert". Klicken Sie mit der rechten Maustaste auf andere Laufwerke und wählen Sie "Statusberichte anpassen". Wählen Sie anschließend "keine Statusberichterstellung". Lassen Sie das Laufwerk C und das Ziellaufwerk die Art und Weise, wie sie sind. Möchten Sie auf dem Laufwerk C und das Ziellaufwerk sollte bereits deaktiviert werden aufgelistet und als "Backup-Ziel".

Task B8 - German MT output:

Vollständiger Systemscan Fehler

Ich vor kurzem installiert Norton Internet Security 2012 auf meinem PC mit Windows 7. Wenn ich Norton und wählen Sie die Option, führen Sie eine vollständige Systemprüfung durch den Scan Fenster (rot): Fehler bei Scan. Dies tritt auf, nachdem das System neu gestartet werden muss. Wenn ich wählen Sie erneut die vollständige Systemprüfung Option ohne Neustart, geschieht nichts (ich nicht den Scan Fenster oder die Fehlermeldung). Wenn ich wählen Sie Quick Scan arbeitet auf "OK". Zuvor hatte ich Norton Internet Security 2011 installiert ist, und ich habe keine Probleme mit der vollständigen Systemprüfung. Vielen Dank für Ihre Hilfe.

Hallo, Willkommen bei Norton Community! Bitte entschuldigen Sie die Unannehmlichkeiten, die Sie mit vollständigen Systemprüfung. Überprüfen Sie im Sicherheitsverlauf, und überprüfen Sie, ob es gibt alle Fehler protokolliert für den vollständigen Systemscan. Sie sind möglicherweise in der Lage, um den Fehler zu finden, Anzahl und Modul von den Informationen, die möglicherweise hilfreich für uns bei der Festlegung der Frage. In der Zwischenzeit können Sie den Computer im abgesicherten Modus neu starten, und doppelklicken Sie auf das Symbol Norton eine vollständige Systemprüfung ausführen. Überprüfen Sie, ob es führt die Scans oder erkennt Bedrohungen. Wenn das Problem weiterhin auftritt, starten Sie den Computer auf den normalen Modus zurück und versuchen Sie es Norton Power Eraser: Http: / / www.norton.com / NPE. Wenn die NPE auch nicht erkennen, haben Sie möglicherweise Probleme zu deinstallieren / installieren Sie Ihr NIS 2012 mit Norton Entfernen & Neuinstallation Tool: Http: / / www.norton.com / nrnr. Lassen Sie uns

wissen, wie es weitergeht. Vielen Dank.

Task B9 - German MT output:

Wiederherstellen Scheitern von Ghost 15

Situation: Ein seltsames Verhalten von Windows 7 prof 64b, (Programme nicht starten, kein Netzwerk Zugriff) nach meinem Mobiltelefon zu entsperren. Dies führte mich Versuch der Wiederherstellung von Laufwerk C: Wiederherstellung boot-CD startet gut aus externen USB-drive aber nach 2 / 3rds der Vollendung es stoppt und Ursachen dieser Fehlermeldung: "Fehler EA390013: Am Ende der etwas". Ich habe versucht, die verschiedenen Backup-Sätze (Backup alle 2 Stunden), mit und ohne die Streichung der Ziellaufwerk. Computer ist ein Sony Vaio Notebook mit doppelten SSD-Laufwerk in RAID 0 Konfiguration. Ich hoffe wirklich, kann jemand geben Sie mir einige gute Ratschläge erteilt. Ich muss mein Computer dringend.

Während der Wiederherstellung der Festplatte gestartet, löschen Sie die Festplatte wiederherstellen. Öffnen Sie die Eingabeaufforderung und befolgen Sie die Anweisungen hier, beginnend mit Schritt 3 fort. Wenn Sie auf # 11, machen Sie sich keine Sorgen über die Schaffung einer neuen Partition als Ghost wiederhergestellt werden. Http: / / www.sevenforums.com / Tutorials / 52129-disk-clean-clean-all-diskpart-command.html. Sicherstellen, DASS SIE DIE WIPING CORRECT DISK UND DASS SIE NICHT WIPE DIE DISK MIT DER IMAGE file ON IT!!! Sobald getan, starten Sie den Computer wieder in die SRD, und versuchen Sie, die Sie wiederherstellen möchten.

Task B10 - German MT output:

USB Flash Drive Wiederherstellung

Hallo, ich habe meinen N360 (ver. 20.2.0.19) Backup auf einem Sandisk CRUZER USB Flash Drive (16 GB). Es gibt auch andere Dateien (überwiegend Fotos) auf die USB Flash-Laufwerke, die nicht im N360 Backup. Wenn ich die Flash-Laufwerk in den Computer ein Backup ausführen, gestern die Flash-Laufwerk wurden keine Dateien ist leer (). Ich glaube nicht, Norton Ghost. Meine 2 Fragen sind: Wie kann ich über

die Wiederherstellung der Dateien aus meiner Flash-Laufwerk (vorausgesetzt, sie sind noch immer dort)? 360 oder Ghost helfen können, dies zu tun? Vielen Dank.

Willkommen, wie sie wiederhergestellt werden können, wird davon abhängen, wie sie fehlt. Hat eine physische oder elektrische Trauma aufgetreten? Haben Sie die Flash-Laufwerk und neu installiert? Vielen Dank.

Task B11 - German MT output:

Wie Sie am besten installieren Sie eine neue Kopie von NIS 2012 über eine vorhandene aber zeitlich begrenzte Installation?

Hallo. Ich habe Norton Internet Security 2011 auf meinem Computer installiert. Ich habe die Vorteile des freien ein Upgrade auf NIS 2012. Dies läuft in wenigen Tagen. Ich habe eine ganz neue Einzelhandelsversion (Produktpaket) von NIS 2012. Was ist die beste Art und Weise ti installieren Sie die neue Lizenz? Sollte ich die aktuellen NIS 2012 deinstalliert, dann installieren Sie die neue Kopie? Ich möchte meine bestehenden Einstellungen zu halten. Ich kann also installieren Sie die neue Kopie über die vorhandene Version? Drittens gibt es eine Möglichkeit, kann ich nur die neuen Produktschlüssel und die vorhandene Installation mit einen längeren Zeitraum? Vielen Dank für jede Hilfe und Beratung.

Hallo, brauchen Sie nicht zu deinstallieren / neu zu installieren. Alles, was Sie tun müssen, ist, geben Sie Ihren neuen Aktivierungsschlüssel in Ihrem aktuellen Abonnement, bevor es abläuft. Nur warten, bis Ihr Abonnement ist ein oder zwei Tage, bevor es abläuft, öffnen Sie NIS und klicken Sie auf Verlängern.

Task B12 - German MT output:

Ich richtig deinstalliert haben, aber die Lizenz wurde nicht deaktiviert

Ich habe das Norton Removal Tool verwendet, entfernen Sie Norton 360 von meinem alten HDD dann installiert ist es ein neues. Aber ich komme ich zu finden, 3 Lizenzen verfügbar, aber jetzt habe ich nur 2. Ich möchte eine andere haben. Auch die Live Chat Unterstützung ist von absolut keine verwenden, weil es nicht einmal den Chat Seite geladen werden.

Wechseln Sie hier, und melden Sie sich mit Ihrem Norton Account https: / / manage.norton.com / Account. Klicken Sie auf der Registerkarte "Mein Account". Dann unter Ihrer Lizenz erhalten Sie die Aktivierungen verwendet und eine "Papierkorb" Symbol können, auf die Sie klicken können, um die Lizenz vom System entfernen.

Appendix F

Post-Editing Guidelines

Tipps zum Bearbeiten:

- Bearbeiten Sie den Text basierend auf Ihrer Interpretation der Bedeutung so, dass er flüssiger und klarer wird.

- Versuchen Sie z.B. die Wortfolge und Rechtschreibung zu korrigieren, wenn diese so unpassend sind, dass der Text nur schwer oder nicht zu verstehen ist.

- Falls Wörter, Satzteile oder Zeichensetzung in dem Text völlig akzeptal sind, versuchen Sie diese unbearbeitet zu verwenden, anstatt sie mit etwas Anderen zu ersetzen.

- Wenn Sie mit Referenz zum Originaltext arbeiten, stellen Sie sicher das keine Informationen versehentlich hinzugefügt oder gelöscht wurden.

Appendix G

Plain Language Statement and Informed Consent - Evaluators

Umfrage (Bewertung)

Einverständniserklärung

In diesem Teil der Umfrage werden die Studie und ihre Aufgaben erklärt. Ihr Einverständnis ist zu allen Teilen erforderlich, damit Sie an dieser Studie teilnehmen können.

Teilnahmeinformationen

Der Titel der Studie ist ,,Gemeinschaftsbasiertes Nachbearbeiten maschinell übersetzter Inhalte" (,,Community-based post-edited machine-translated content"). Diese Studie wird im Rahmen des Fachbereiches der Angewandten Sprache und Interkulturellen Wissenschaften (School of Applied Language and Intercultural Studies) in der Fakultät der Geisteswissenschaften und Sozialwissenschaften durchgeführt. Linda Mitchell ist die leitende Forschungswissenschaftlerin dieser Studie, die über Email linda.mitchell7@mail.dcu.ie oder Telefon +353858122880 erreicht werden kann. Weitere an diesem Experiment beteiligte Wissenschaftler sind Dr. Sharon O'Brien, Dozentin an der Dublin City University, Dr. Fred Hollowood, Leiter der sprachtechnischen Forschung bei Symantec und Dr. Johann Roturier, leitender Forschungsingenieur bei Symantec. Diese Forschungsstudie wird als Teil des ACCEPT-Projekts (Bewilligungnummer: 288769) von der EU-Kommission subventioniert.

Sie werden in dieser Studie über einen Zeitraum von einer Wochen dreizehn Aufgaben bewerten, die maschinell übersetzte Inhalte, die von Mitgliedern der Norton Community korrigiert wurden, beinhalten. Davor werden Sie gebeten eine kurze Umfrage auszufüllen. Insgesamt sollte dies nicht mehr als sechs Stunden in Anspruch nehmen.

Die Studie ermöglicht es Ihnen, einen Einblick in die Norton Community und in den Satus Quo maschineller Übersetzung zu gewinnen. Des Weiteren wird dies der Norton Community langfristig gesehen zugute kommen, da sich die Möglichkeit bietet, die Inhalte der Community zu bereichern und schnelleren Zugriff zu Lösungsansätzen zu erlangen. Die Ergebnisse dieser Studie werden nach Vollendung der Doktorarbeit (Ende 2014) auf folgender Website öffentlich zugänglich sein: http://doras.dcu.ie/.

Die im Zuge der Studie erfassten Daten werden entsprechend den Vorschriften, die von der EU-Kommission für Forschungsprojekte aufgesetzt wurden, gesammelt, gespeichert und vernichtet. Die Teilnahme an dieser Forschungsstudie ist freiwillig. Jeder Teilnehmer kann die Studie zu jeder Zeit ohne negative Folgen, abbrechen.

Falls Teilnehmer Bedenken bezüglich dieser Studie haben sollten und Kontakt zu einer unabhängigen Person aufnehmen wollen, kontaktieren Sie bitte:
The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel +35317008000

1. Ich habe die zusammenfassende Erklärung gelesen.
2. Ich verstehe die Informationen, die mir präsentiert wurden.
3. Ich hatte die Möglichkeit Fragen zu stellen und diese Studie zu besprechen.
4. Ich habe befriedigende Antworten auf alle meine Fragen erhalten.
5. Ich bin mir darüber im Klaren, dass ich maschinell übersetzte, korrigierte Aufgaben bewerten werde.
6. Ich bin mir darüber im Klaren, dass ich eine Umfrage ausfüllen werde.
7. Ich bin mir darüber im Klaren, dass diese Bewertungen im Rahmen des ACCEPT-Projekts benutzt werden.

✱1. Ich stimme den oben aufgeführten Aussagen zu:

○ ja

○ nein

Einverständiniserklärung

Meine Teilnahme an dieser Studie ist freiwillig. Ich kann die Forschungsstudie zu jeder Zeit und ohne jegliche negative Folgen abbrechen. \newline
Die Daten, die mit meiner Hilfe in dieser Studie erfasst werden, werden entsprechend der Vorschriften, die von der EU-Kommission für Forschungsprojekte aufgesetzt wurden, erfasst, gespeichert und vernichtet. Ich bin mir darüber im Klaren, dass ich auf die Ergebnisse dieser Studie auf http://doras.dcu.ie/ nach Vollendung der Doktorarbeit (Ende 2014) zugreifen kann.

Ich habe die Informationen gelesen und verstanden. Meine Fragen und Bedenken wurden von den Forschern beantwortet, und ich habe eine Kopie dieser Erklärung erhalten.

Umfrage (Bewertung)

G-3

✳2. Ich stimme den Bedingungen dieser Studie zu:

○ ja

○ nein

Survey (Evaluation)

Informed Consent

This part of the survey presents the study and the tasks involved. Your consent is required for all parts in order for you to take part in the study.

Plain Language Statement

The title of this study is "Community-based post-editing of machine translated content". This experiment is conducted within the School of Applied Language and Intercultural Studies (SALIS) in the faculty of humanities and social sciences. The principal investigator is Linda Mitchell. Other investigators involved in this experiment are Dr. Sharon O'Brien, senior lecturer at DCU, Dr. Fred Hollowood, Director for Language Technologies Research in Symantec and Dr. Johann Roturier, Principal Research Engineer in Symantec. This research study is funded by the European Commission as part of the ACCEPT project (Grant Agreement Number: 288769).

You will evaluate thirteen tasks containing content that has been machine translated and corrected by members of the Norton community over a duration of one week. Prior to this, you will be asked to fill in a short survey. In total, this should not take longer than six hours.

Taking part in this research study will provide you with the opportunity to learn more about the current state of machine translation. Furthermore, it will help the Norton Community in the long run, as the forum content is going to be enriched and improved. The results of this study will be publicly available on http://doras.dcu.ie, once the PhD thesis is completed (late 2014).

The data that is collected with my help during this study will be collected, stored and destroyed in compliance with the regulations for research projects set up by the European Commission. My involvement in this study is voluntary. I may withdraw from the Research Study at any point without any negative consequences.

If participants have concerns about this study and wish to contact an independent person, please contact:
The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Phone 01-7008000

1. I have read the Plain Language Statement (or had it read to me)
2. I understand the information provided
3. I have had an opportunity to ask questions and discuss this study
4. I have received satisfactory answers to all my questions
5. I am aware that I will be required to evaluate content
6. I am aware that I will complete a survey
7. I am aware that the evaluations will be used within the framework of the ACCEPT project

*1. I agree with the statements presented above:

○ yes

○ no

Informed Consent

My involvement in this study is voluntary. I may withdraw from the Research Study at any point without any negative consequences.

The data that is collected with my help during this study will be collected, stored and destroyed in compliance with the regulations for research projects set up by the European Commission.
I am aware that I can access the results of this study on http://doras.dcu.ie/, once the PhD thesis is completed (late 2014).

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form.

*2. Therefore, I consent to take part in this research project:

○ yes

○ no

Appendix H

Evaluation Guidelines

German guidelines:

Hintergrund der Studie:

Der Inhalt, der in dieser Auswertung verwendet wird, besteht aus Fragen und Antworten aus der Norton Community (US). Für jede deutsche Quellübersetzung gibt es fünf deutsche Übersetzungsvorschläge. Diese sind maschinell übersetzte Sätze, die von Mitgliedern der Norton Community korrigiert wurden.

Zur Bewertung:

Das Ziel dieser Auswertung ist es, verschiedene Übersetzungsvorschläge mit einer Quellübersetzung zu vergleichen. Dies beinhaltet zum einen zu ermitteln, wie viel von der Bedeutung der Quellübersetzung in den gezeigten Übersetzungsvorschlägen enthalten ist, und zum anderen, die sprachliche Qualität der Übersetzungsvorschläge einzuschätzen. Bitte lesen sie die Quellübersetzung und die fünf Übersetzungsvorschläge. Danach bewerten Sie die Vollständigkeit und sprachliche Qualität (bzw. Sprachfluss) auf den in den Aufgaben angegebenen Skalen.

Hinweise:

- Die Vollständigkeit und die sprachliche Qualität (bzw. Sprachfluss) sollten mit steigender Anzahl der Fehler abnehmen.

- ABER: Selbst wenn nur die Bedeutung eines einzigen Wortes falsch ist und die Bedeutung des gesamten Satzes umkehrt, sollte die Vollständigkeit stark abnehmen.

- Fehlende Informationen sollten mit Punktabzug in Abhängigkeit von der Menge bestraft werden.

- Zusätzliche Informationen sollten in Abhängigkeit davon, wie sehr sie die Bedeutung der Quellübersetzung verändern, bestraft werden.

- Auf Ihrem Fachwissen basierend, sollten Sie Variationen in den Übersetzungen, falls die Bedeutung des Satzes noch immer korrekt ist, tolerieren.

- Das Übernehmen eines englischen Begriffes oder die Übersetzung ins Deutsche sollte ebenso toleriert werden.

- Verwenden Sie nicht mehr als ein paar Minuten zur Bewertung eines Satzes und dessen Übersetzungsvorschläge.

- Schließen Sie eine Aufgabe, wenn Sie gerade nicht an ihr arbeiten. (Die Zeit für die Bewertung wird aufgezeichnet.)

Vielen Dank fürs Mitmachen und viel Spaß bei der Bewertung!

English translation:

Background of the Study:

The content that is used in this evaluation consists of questions and answers from the Norton Community (US). For each of the German reference translations there are translation candidates. These are machine translated sentences, which have been corrected by members of the Norton Community.

About the Evaluation:

The aim of this evaluation is to compare several translation candidates to a reference translation. This involves, firstly to identify how much meaning of the reference translation is also expressed in the translation candidates presented and secondly, to rate the fluency of the translation candidates. Please read the reference translation and the five translation candidates. Subsequently, rate the fidelity and fluency using the scales that are presented in the tasks.

Guidelines:

- Adequacy and fluency should decrease with an increasing number of mistakes.

- BUT: if the meaning of just one word is wrong and this leads to a meaning opposite to the intended being attributed to the whole sentence, then adequacy should decrease dramatically.

- Missing information should be penalised according to quantity.

- Additional information should be penalised according to how much the meaning of the reference translation is changed.

- Based on your domain knowledge, variation in the translations should be tolerated if the meaning of the sentence is still correct.

- Adopting an English term in German, as well as the translation of the same term should be equally tolerated.

- Do not spend more than a few minutes on evaluating a sentence and its corresponding translation candidates.

- Close a task if you're currently not working on it. (The time of the evaluation will be recorded.)

Thank you very much for your participation and enjoy the evaluation!

Appendix I

Fifty Segments for Community Evaluation

1   Anschließend wählen Sie den Link "Exportergebnisse", um die Prüfung der unten angezeigten Ergebnisse in eine Textdatei zu übernehmen.

2   5) Klicken Sie auf "Deaktivieren" für diese Datei und starten Sie den Chrome Browser neu. Abmelden und neu gestartet haben keine Auswirkungen.

3   Abschließend möchte ich das Problem isoliert und festgestellt, dass die NU war Löschen von Elementen in Bezug auf die Toolbox sowohl in AppData und in der Registrierung.

4   Alle meine Produkte auf dem neuesten Stand sind.

5   Alles, was Sie tun müssen, ist, Ihren neuen Aktivierungsschlüssel in Ihrem aktuellen Abonnement einzugeben, bevor es abläuft.

6   Bleiben Sie gut und surfen Sie sicher

7   Der lokale Computer hat unter Umständen nicht die erforderlichen Registrierungsinformationen oder DLL Dateien geforderten Berichte anzeigen, Meldungen von einem Remote-Computer."

8   Die einzige Aufzeichnung, die ich ist der Verlauf meines Kaufs in meinem Android Markt und eine Kopie der Rechnung.

9   Die einzige Lösung für mich besteht darin, das Programm zu deinstallieren und erneut zu installieren.

10 Die E-Mail war gehackt.

11 Dies geschieht, denn es ist aktiviert, wenn es keine Verbindung mit dem Produktschlüssel mit Pack enthalten.

12 Dies hilft Ihnen mit Ihrer Entscheidung, ob sollten Sie jedes der aufgeführten Websites blockieren.

13 Dies läuft in wenigen Tagen aus.

14 Dieses Problem besteht seit ein paar Tagen, bis dahin hat es einwandfrei funktioniert.

15 Fehler EA390013: Am Ende der etwas.

16 Gibt es irgendeine Möglichkeit, die gleiche Lizenz auf meinen neuen Laptop zu übertragen, sodass ich mein Abonnement in vollem Umfang nutzen kann?

17 Hallo, die beste Option wäre, wenn Sie dies mit dem Kundensupport besprechen.

18 Hallo, My Mac wurde mit W32.Downadup.B infiziert.

19 Hallo, scheint das Kunststück getan haben.

20 Ich hatte deaktiviert den Dienst völlig für eine Woche, würde ich dachte, und versuchen Sie es erneut heute – und es funktioniert.

21 Hallo, sind Sie sicher, dass Norton TeamViewer blockiert?

22 Hallo, Willkommen in der Norton Community!

23 Ich also über Datenträger installieren müssen, um zu sehen, ob jemand muss ich mich mit diesem helfen können.

24 Ich habe die Norton-Erweiterung in den Chrome-Bericht Einstellungen deaktiviert und das Problem geht weg.

25 Ich habe eine vollständige Systemprüfung durchgeführt und nur 4 Cookies erkannt wurden.

26 Ich habe versucht, mit dem Startmenü deinstallieren und nach, dass die Verwendung des $Norton_Removal_Tool$.

27 Ich habe, die Datei zu schützen, meine geschützten Dateien nicht öffnen, weil er sich selbst mit der richtige Name und Kennwort des Administrators.

28 Ich hatte diese Website für einen kurzen Zeitraum ruhen, bis ich die ID Safe wouldn 't ausfüllen zu unterzeichnen.

29 Ich höre wie das Laufwerk arbeitet, aber nichts wird geöffnet.

30 Ich kann nicht den Aktivierungscode Aktivierung über das norton.com / nms Link.

31 Ich machte ein Upgrade zu Version 13.

32 IE9 und NIS Symbolleiste Frage

33 In der Vergangenheit, wenn ich mit Norton NIS 2012 in den Zwischenspeicher kopiert habe, waren die Tracking-Cookies aufgelistet, wenn ich sie in Notepad eingefügt habe.

34 In diesem Fall schlage ich vor der Registrierung auf einem der Malware entfernen.

35 In fast allen Fällen verursacht werden, die Ereignisse von legitimen Programmen oder Windows Prozesse, und nichts zu befürchten.

36 Ist dies ein Virus und wenn ja, wie kann ich es loszuwerden.

37 Jede Ressource sagt, dass diese Infektion ist nur für PCs und bietet Lösungen auf PCs, um es zu lösen.

38 Jetzt mit der neuen Version nach Re- Abonnements, die nicht sagen NIS 2013 nur sagt (Norton Internet Security) ver. 20.2.1.22.

39 Nach der Neuinstallation funktioniert alles gleich, keine spezifischen Tracking-Cookie Informationen beim einfügen.

40 Norton wird automatisch neue Regeln für das Programm konfigurieren.

41 Sie hat auf diesem keinen Virenschutz.

42 Sie hatte, dass ein Hacker.

43 Sie sollten nun , die detaillierten Tracking-Cookie Informationen sehen können.

44 Soweit ich weiß können Sie Websites nicht mit Firefox oder Chrome anpinnen.

45 Um das Problem zu arbeiten, müssen Sie einen Quick Scan ausführen.

46 Vielen Dank für helfen.

47 Wählen Sie den Link "Exportergebnisse", um die Prüfung der unten angezeigten Ergebnisse in einer Textdatei:

48 Was ist der beste Weg um NIS 2012 zu installieren. über eine exentierende Version?

49 Wenn ich wählen Sie Quick Scan arbeitet auf "OK".

50 Wenn Sie herunterladen und installieren konnten die zweite Programm für das Ergebnis verringert würde, Schutz.

Appendix J

Domain Specialist Evaluation Split per Post-Editor

Table J.1 and J.2 display all data points comparing the raw MT output and the PE results for group A and B with * as statistically significant and ** as highly statistically significant. Particularly low agreement without statistical significance is marked in grey in Table J.2.

	MT fidelity	PE fidelity	difference	MT fluency	PE fluency	difference
E1 - A1	3.18	3.85	0.67**	2.69	3.63	0.94**
E2 - A1	3.11	3.39	0.28**	2.54	3.41	0.88**
E3 - A1	3.28	3.84	0.56*	2.78	3.76	0.98**
E1 - A2	2.85	4.35	1.50**	2.45	3.95	1.50**
E2 - A2	2.99	4.16	1.17**	2.43	3.63	1.20**
E3 - A2	2.94	4.50	1.56**	2.50	4.17	1.67**
E1 - A3	3.10	4.42	1.34**	2.71	4.22	1.51**
E2 - A3	2.91	4.00	1.10**	2.39	4.13	1.74**
E3 - A3	2.98	4.52	1.55**	2.58	4.52	1.94**
E1 - A4	3.07	4.00	0.93**	2.72	3.65	0.93**
E2 - A4	3.00	3.99	0.99**	2.62	3.75	1.13**
E3 - A4	3.00	4.08	1.08**	2.65	3.98	1.33**
E1 - A5	3.07	4.28	1.21**	2.51	3.89	1.38**
E2 - A5	2.89	4.39	1.50**	2.46	3.81	1.35**
E3 - A5	2.90	4.71	1.81**	2.58	4.40	1.81**
E1 - A6	3.12	4.59	1.47**	2.51	4.30	1.79**
E2 - A6	2.94	4.29	1.35**	2.33	4.28	1.95**
E3 - A6	3.04	4.67	1.63**	2.32	4.70	2.39**
E1 - A7	2.67	4.25	1.58**	2.36	4.02	1.66**
E2 - A7	2.78	4.30	1.52**	2.27	3.91	1.64**
E3 - A7	2.66	4.59	1.93**	2.27	4.51	2.23**
E1 - A8	3.06	4.29	1.22**	2.57	3.92	1.35**
E2 - A8	2.90	4.13	1.23**	2.32	3.81	1.49**
E3 - A8	3.07	4.47	1.40**	2.55	4.33	1.78**
E1 - A9	3.04	4.61	1.57**	2.60	4.17	1.58**
E2 - A9	2.97	4.15	1.18**	2.42	3.88	1.46**
E3 - A9	2.94	4.63	1.69**	2.46	4.51	2.01**
mean	2.98	4.28	1.30	2.50	4.05	1.54

TABLE J.1: Comparison between raw MT output and PE results - Full Data Group A

	MT fidelity	PE fidelity	difference	MT fluency	PE fluency	difference
E4 - B1	4.62	4.74	0.12*	2.04	3.30	1.26**
E5 - B1	3.42	4.52	1.10**	2.76	3.68	0.91**
E6 - B1	4.00	4.72	0.72**	2.69	3.67	0.98**
E7 - B1	4.29	4.55	0.26**	2.60	3.18	0.58**
E4 - B2	4.57	4.76	0.2*	2.39	3.86	1.47**
E5 - B2	3.58	4.77	1.2**	3.12	4.45	1.33**
E6 - B2	4.01	4.80	0.79**	2.75	4.49	1.74**
E7 - B2	4.37	4.71	0.34**	2.76	3.97	1.21**
E4 - B3	4.71	4.82	0.11*	2.35	3.67	1.31**
E5 - B3	3.81	4.44	0.63**	3.04	3.95	0.90**
E6 - B3	4.35	4.72	0.37**	2.96	4.02	1.06**
E7 - B3	4.46	4.70	0.24**	2.68	3.58	0.90**
E4 - B4	4.65	4.83	0.18**	2.22	4.19	1.97**
E5 - B4	3.45	4.82	1.37**	2.91	4.47	1.56**
E6 - B4	4.01	4.89	0.88**	2.81	4.56	1.75**
E7 - B4	4.29	4.82	0.53**	2.55	4.13	1.58**
E4 - B5	4.59	4.79	0.20*	2.16	3.96	1.80**
E5 - B5	3.54	4.77	1.23**	3.05	4.39	1.33**
E6 - B5	3.98	4.86	0.89**	2.85	4.49	1.64**
E7 - B5	4.35	4.77	0.42**	2.72	3.92	1.19**
E4 - B6	4.65	4.85	0.19*	2.23	4.15	1.92**
E5 - B6	3.60	4.73	1.13**	2.88	4.46	1.58**
E6 - B6	4.01	4.79	0.78**	2.56	4.47	1.91**
E7 - B6	4.45	4.83	0.38**	2.51	4.12	1.60**
E4 - B7	4.59	4.70	0.11*	1.97	2.90	0.93**
E5 - B7	3.39	4.16	0.77**	2.80	3.50	0.70**
E6 - B7	4.04	4.40	0.36**	2.61	3.44	0.83**
E7 - B7	4.36	4.56	0.20**	2.66	3.20	0.54**
E4 - B8	4.59	4.47	0.13	1.90	2.85	0.95**
E5 - B8	3.52	3.77	0.25*	2.76	3.58	0.82**
E6 - B8	4.01	4.03	0.02	2.60	3.55	0.94**
E7 - B8	4.24	4.39	0.15*	2.47	3.25	0.78**
E4 - B9	4.66	4.71	0.05	2.11	3.41	1.30**
E5 - B9	3.44	4.58	1.14**	2.76	4.16	1.40**
E6 - B9	3.80	4.58	0.78**	2.56	4.32	1.76**
E7 - B9	4.17	4.65	0.47**	2.56	3.76	1.20**
mean	4.13	4.64	0.52	2.59	3.86	1.27

TABLE J.2: Comparison between raw MT output and PE results - Full Data Group
B

Appendix K

High-Scoring Segments

1   1) Öffnen Sie Chrome und geben Sie "about:plugins" in die Adressleiste ein.

2   5) Klicken Sie auf "Deaktivieren" für diese Datei und starten Sie den Chrome Browser neu.

3   Aber NU hätte diese nicht löschen dürfen.

4   Abmeldung und Neustart zeigen keine Wirkung.

5   Alle meine Produkte sind auf dem neuesten Stand .

6   Auch die Einzelhandelsversion von NMS wird nicht in Ihrem Norton Account als Norton-Produkt angezeigt.

7   Beenden Sie Norton und starten Sie TeamViewer neu.

8   Beide PCs laufen mit Windows 7 und Norton 360.

9   Der PC ist immer eingeschaltet und automatische Updates sind aktiviert.

10   Die einzige Lösung für mich besteht darin, das Programm zu deinstallieren und erneut zu installieren.

11   Die einzige Lösung für mich besteht darin, das Programm zu deinstallieren und neu zu installieren.

12   Dies hilft Ihnen mit Ihrer Entscheidung, welche der aufgeführten Websites sie blockieren möchten.

13   Dieses Problem besteht seit ein paar Tagen, bis dahin hat es einwandfrei funktioniert.

14   Eine Freundin von mir hat Trojaner auf ihrem Computer.

15   File Guard öffnet Dateien sogar mit dem richtigen Administratorkennwort nicht.

16   Hallo, sind Sie sicher, dass Norton TeamViewer blockiert?

17   Hallo, vielen Dank für die Aktualisierung.

18   Hallo, mein Mac wurde mit W32.Downadup.B infiziert.

19   Hallo.

20   Ich habe drei Prüfungen mit NIS und eine vollständige Systemprüfung mit der kostenlosen Malwarebytes Anti-Malware-Software durchgeführt.

21   Ich habe versucht, per Startmenü zu deinstallieren und danach auch unter Verwendung des Norton _ Removal _ Tool.

22   Ich hoffe, dass dies hilfreich ist.

23   Ich muss den Dateischutz ausschalten, um an meine geschützten Dateien heranzukommen, weil sie sich selbst mit dem richtigen Namen und Kennwort des Administrators nicht öffnen lassen.

24   Ich muss diese Infektion beseitigen, da sie meinen Browser und andere Computer in meinem Netzwerk verlangsamt.

25   Ist das jetzt in ver. 20.2.1.22 normal?

26   Ist das jetzt normal mit der ver. 20.2.1.22?

27   Jedoch hätte NU diese nicht löschen sollen.

28   Letztlich habe ich das Problem isoliert und festgestellt, dass die NU sowohl in AppData als auch in der Registrierung Elemente in Bezug auf die Toolbox gelöscht hat.

29   Nun sollten Sie die detaillierten Tracking-Cookie Informationen sehen können.

30   Öffnen Sie diese neu gespeicherte Textdatei .

31   Seiten frieren beim Laden ein, ich kann überhaupt nicht scrollen oder mit der Seite interagieren.

32   Sie hat keinen Virenschutz auf dem Computer.

33   Sie können jedoch den Eintrag TeamViewer aus der Liste entfernen, um eine mögliche Fehlkonfiguration automatisch zu reparieren.

34   Sie sollten nun die detaillierten Tracking-Cookie Informationen sehen.

35   Stellen Sie zuerst sicher, dass TeamViewer nicht ausgeführt wird.

36   Um das Problem zu umgehen, müssen Sie einen Quick Scan ausführen.

37   Verwenden Sie diesen Namen, wenn File Guard Sie nach dem Administrator-passwort fragt.

38   Vielen Dank

39   Vielen Dank für die Hilfe.

40   Vielen Dank für Ihre Hilfe.

41   Vielen Dank!

42   Vielen Dank.

43   Vielleicht ist es ein bisschen zu aggressiv.

44   Was passiert, wenn eine NAV Lizenz abläuft?

45   Weitere Informationen finden Sie in der Hilfe.

46   Wenn ich also versuche, ein Dokument zu scannen, kann der Scanner nicht gefunden werden.

47   Wie heißt Norton?

48   Zu Ihrer Information, das NMS-Upgrade wird nicht in Ihrer Norton Account-Produktliste erscheinen, wenn Sie das Upgrade über Android Market erworben haben.

Appendix L

Low-Scoring Segments

1   "Fehler EA390013. Oder ähnlichem.

2   1) Öffnen Sie Chrome und geben Sie "über: Plug-ins" in die Adressleiste ein.

3   4) Suchen Sie die Datei mit dem Standort, endet in Pepper Flash oder pep-flashplayer.dll.

4   Aber Ghost IMMER sagt "Gefährdet" (in der Systemleiste und Ghost Home Bildschirm), so ich Ghost immer öffnen soll es manuell überzuprüfen...

5   Aber Ghost sagt IMMER "Gefährdet" (in der Systemleiste und Ghost Home Bildschirm), damit muss ich Ghost immer öffnen, um es manuell zu aktivieren...

6   Aber Ghost sagt IMMER "Gefährdet" (in der Systemleiste und Ghost Home Bildschirm), so ich Ghost immer öffnen soll, um manuell zu aktivieren...

7   Aber ich bin gekommen, weil ich gefunden habe, dass ich 3 Lizenzen zur Verfügung habe, aber jetzt habe ich nur 2.

8   Aber ich komme ich zu finden, 3 Lizenzen verfügbar, aber jetzt habe ich nur 2.

9   Aber NU sollte nicht gelöscht werden.

10 Abschließend möchte ich das Problem isoliert und festgestellt, dass die NU war Löschen von Elementen in Bezug auf die Toolbox sowohl in AppData und in der Registrierung.

11 Adguard NIS2012 und

12 Alle Informationen über das wäre sehr geschätzt werden.

13 Alles, was Sie tun müssen, geben Sie Ihren neuen Aktivierungsschlüssel in Ihr aktuelles Abonnement, bevor es abläuft.

14 Alles, was Sie tun müssen, geben Sie Ihren neuen Aktivierungsschlüssel in Ihr aktuelles Abonnement, bevor es abläuft.

15 Also starten Sie die Installation manuell, anstatt die Autoausführung aktivieren und deaktivieren.

16 Anzeigen eines Image in einer Spam-Nachricht wird es dem Spammer ermöglichen zu wissen, dass Ihre E-Mail-Adresse "aktiv" und bereit für eine schwere Bombardierung ist.

17 Anzeigen eines Image in einer Spam-Nachricht wird es ermöglichen, den Spammer zu erkennen, dass Ihre E-Mail-Adresse "lebt" und ist bereit für schwere Bombardierung.

18 Auch alle Browser ist sehr langsam.

19 Auch die Live Chat Unterstützung ist von absolut keine verwenden, weil es nicht einmal den Chat Seite geladen werden.

20 Auch eine Deaktivierung der Module: anti-spam und eingehendes untersruchen blockiert die Bilder.

21 Auf dem erweiterten Bildschirm sehe ich mein Laufwerk C: mit einem grünen Kreis und 2 anderen C: Laufwerke mit Fragezeichen.

22 Auf den PC war immer das automatische Update aktiviert.

23 Aus Sicherheits Gründe, ist es am besten, dass das deaktiviert ist.

24 Beachten Sie die folgenden Microsoft Artikel: Bilder, die Windows Mail blockiert

25 Beim Laden der Seite Abstürze blättern, kann ich nicht auf alle oder Interaktionen mit Seite.

26 Beim öffnen einer Seite, IE öffnet eine Isolierte Webseite, so meldet sich nicht die Toolbar.

27 Benötigen HILFE SCHNELL!!

28 Benötigen SIE HILFE SCHNELLE!!

29 Bereinigung im Gang für 2-1 / 2 Tage... Hilfe!

30 Bereinigung läuft für 2-1 / 2 Tage... Hilfe!

31 Bin ich übrigens mit hilfe dieser Funktion unter dem Dach der NIS?

32 Bitte entschuldigen Sie für die Unannehmlichkeiten Sie mit vollständigen Systemprüfung hat.

33 Bitte überprüfen Sie im Sicherheitsverlauf, ob es gibt einige eingetragene Fehler für den vollständigen Systemscan.

34 Bleiben Sie gut und sicher im Internet.

35 Bleiben Sie gut und surfen Sie sicher

36 Dachten vielleicht war die WLAN-Verbindung.

37 Dann unter Ihrer Lizenz erhalten Sie die Aktivierungen verwendet und eine "Papierkorb" Symbol können, auf die Sie klicken können, um die Lizenz vom System entfernen.

38 Dann unter Ihrer Lizenz können Sie die verwendete Aktivierungen und eine "Papierkorb" Symbol sehen, auf man kann klicken die Lizenz aus diesem System zu entfernen.

39 Dann sind Sie auf sichere Seite.

40 Das Anzeigen eines Bildes in einer Spam-Nachricht ermöglicht es dem Spammer zu wissen, dass Ihre E-Mail-Adresse "live" ist und bereit für schwere Bombardierung.

41 Das ist kein Aktivierungsschlüssel für eine Binnenmarkt-Version von NMS.

42 Das Live Chat Kundensupport ist absolut keine Verwendung, weil es einmal den Chat Seite nicht geladen werdet.

43 Denken Sie daran, wenn es mit jedem anderen Sicherheitsprogramm installiert kommst, man soll es gründlich mit dem Programm's Entfernungsprogramm entfernen.

44 Der Datenträger, um NIS 2012 zu installieren, startet nicht automatisch.

45 Der einzige Beweis meines Kaufes ist durch die Kauf Historie meinem Androide Markt Konto und eine Kopie der Rechnung.

46 Der schlimmste Fall wäre, dass sie tatsächlich abbrechen und Sie haben keinen Schutz.

47 Die Backups auf einer externen WD RAID Einheit, ich kann mich nicht an das Modell erinnern, aber es hat zwei 2 TB Laufwerke in einer gespiegelten Array.

48 Die einzige Aufzeichnung, die ich ist der Verlauf meines Kaufs in meinem Android Markt und eine Kopie der Rechnung.

49 Die einzige Lösung für mich ist, das Programm zu installieren und es erneut zu installieren.

50 Die E-Mail war gehackt.

51 Die Tracking Cookie Informationen aus dem Einfügen

52 Die Tracking-Cookie Informationen aus den der Paste-Funktion ist weggelassen.

53 Die Tracking-Cookie Informationen einfügen.

54 Die "Anzeigen" Dropdown-Liste in allen "File Insight" Fenstern wurde neugestaltet um die Windows 8 Touch Funktionalität zu unterstützen.

55 Dies deshalb, weil die beiden Programme miteinander kämpfen, was zu tun ist bei verdächtigen Aktionen.

56 Dies führte mich dazu die Wiederherstellung von Laufwerk C: zu versuchen.

57 Wiederherstellung boot-CD startet.

58 Aber vom externen USB-drive nach 2 / 3 der Vollendung stoppt es und zeigt diese Fehlermeldung:

59 Dies führte mich zum Versuch der Wiederherstellung von Laufwerk C:.

60 Wiederherstellen Process startet gut mit einem boot-CD von externen USB-drive aber nach 2/3 der Vollendung es stoppt und ursacht diese Fehlermeldung:

61 Dies geschieht, weil keine Verbindung mit dem Produktschlüssel der im Paket enthalten ist bei der Aktivierung besteht.

62 Dies geschieht, weil wenn es mit dem Produktschlüssel im Pack aktiviert wird, keine Verbindung hergestellt wird.

63 Dies hilft Ihnen mit Ihrer Entscheidung, ob jede der aufgeführten Websites blockieren sollten.

64 Dies kann passieren, wenn es keine Verbindung mit dem Produktschlüssel die enthalten war möglich ist.

65 Dies passiert, weil die beiden darüber kämpfen, was zu tun ist, was auch immer sie verdächtiges gefunden haben.

66 Dies passiert, weil die beiden kämpfen, das zu tun und damit das, was sie gefunden haben, ist verdächtig geworden.

67 Dieses Backup-Profil ist für mein (primärer) Geschäftscomputer und es ist mein Offsite-Backup.

68 Ein einzige Echtzeit Scanner arbeiten muss.

69 Eine andedre Möglichkeit ist, sich in den Norton Account einzuloggen, hier auf den Reiter - mein Account.

70 Eine unbeabsichtigte Folge dieser Konzeptionsänderung, funktioniert das "in Zwischenablage kopieren" in diesem Dialogfeld nicht mehr wie erwartet.

71 Eröffnung einer anhefteten Site an der Taskleiste wird eine isolierte IE Seite öffnen und wird die Norton-Symbolleiste nicht angezeigt.

72 Es gibt keine Hinweise auf irgendetwas bösartigem in dem was Sie veröffentlicht haben.

73 Es gibt keinen Aktivierungs Produktschlüssel für den Marktversion von NMS.

74 Es wird 100% der gleichen funktionieren.

75 Falls es ihnen möglich gewesen wäre das zweite Programm herunterladen und zu installieren, wäre das minimalste Resultat ein geringerer Schutz

76 Fast alle E-Mails mit Bilder in der HTML-Code werden blockiert, und ich muss sie jedes Mal manuell zu sehen zustimmen.

77 Fast alle E-Mails mit Bildern im HTML-Code werden blockiert und ich muss manuell "OK" jedes Mal, sie zu sehen.

78   Fehler EA390013: Am Ende von etwas.

79   Fehler EA390013: Am Ende von irgendwas.

80   Folgen Sie den Anweisungen und sie sollten bereinigt werden.

81   Folgen Sie den Anweisungen und sie sollten sich bereinigt.

82   Ghost 15 Wiederherstellung Scheitern

83   Gibt es eine Möglichkeit erhalte ich mein Benutzername und Kennwort eingegeben werden von Identity Safe?

84   Gibt es eine Möglichkeit, dass ich meinen Benutzernamen und das Passwort bekomme, um mich einzuloggen bei Identity Safe?

85   Gibt es etwas, weiter ich tun sollten, um sicherzustellen, dass ein Hacker nicht eine Datei auf meinem System geladen hat?

86   Guten Abend, ich muss einige helfen.

87   Haben Sie das Flash-Laufwerk neu installiert einen Neustart durchgeführt?

88   Haben Sie die Flash-Laufwerk installiert dann neugestartet?

89   Hallo, brauchen Sie nicht zu deinstallieren / neu zu installieren.

90   Hallo, Einträge in Norton Produktmanipulationsschutz Datensatz zeigen jeden Versuch einer externen Agent für den Zugriff auf einen Norton Datei oder Prozess.

91   Hallo, ich habe ein Problem mit der Norton Erweiterungs Symbolleiste und dem Verlangsamen und Einfrieren beim Laden von Webseiten .

92   Hallo, ich habe jetzt Norton Online Family. installiert.

93   Hallo, ich habe Norton 360 (Version 20.2.0.19) installiert und möchte auf meinem Scandisk Cruzer / USB Flash Drive (16 GB).

94   Hallo, ich verwende mein N360 (ver. 20.2.0.19) Backup auf einem Sandisk CRUZER USB Flash Drive (16 GB).

95   Hallo, Lesung über das Programm auf ihrer Website zu haben scheint es eine aktive Scanner.

96   Hallo, My Mac wurde mit W32.Downadup.B infiziert.

97   Hallo, My Mac wurde infiziert mit W32.Downadup.B.

98   Hallo, scheint Sie, beim Lesen über das Programm auf ihrer Web site, einen aktiven Scanner zu haben.

99   Hier sind meißt Fotos und ein Norton 360 Backup.

100   Http: / / www.bleepingcomputer.com / http: / / www.geekstogo.com / Forum / http: / / www.cybertechhelp.com / forums / http: / / support.emsisoft.com / Forum / 6-help-my-pc-is-infected / http: / / forums.whatthetech.com /.

101 Ich benötige beides so schnell wie möglich bitte!!!

102 Ich benötige bitte beides so schnell wie möglich!!!

103 Ich benötige jemanden von der Seite Symantecs, um die Bereinigungsprozess auf einem meiner 2 mit NOBU verbunden Computer zu Ende zu bekommen.

104 Ich bin angemeldet über das Admin-Konto und lege den Datenträger in das CD-Laufwerk ein.

105 Ich bin zögerlich zur Verwendung des Norton Power Eraser Tool.

106 Ich dachte an die die WLAN-Verbindung.

107 Ich denke, die Datei wird mit einem Versuch erfolgreich entfernt werden (nach Erfolg würde der Befehl rm zeigen "Datei nicht gefunden").

108 Ich denke, Sie entfernen die Datei mit einem Versuch erfolgreich (bei Erfolg würde der Befehl rm zeigen "Datei nicht gefunden").

109 Ich glaube nicht, dass der Fehler bei Norton Ghost liegt.

110 Ich habe als Admin-Konto angemeldet und ich habe die Datenträger in das CD-Laufwerk eingelegt.

111 Ich habe auf 360you vor über 3 Monaten aktualisiert und bis vor 7-10 Tagen nicht ein Problem gehabt.

112 Ich habe das Norton Removal Tool verwendet,Sie Norton 360 von meinem alten HDD zu entfernen und dann das neue Prgroamm installiert.

113 Ich habe keine Woche Zeit für den Vorgang.

114 Ich habe NMS in einem Markt in das Ende Jan. gekauft.

115 Ich habe nur ein einzelnes Laufwerk C: und eine Backupaufgabe die immer nach dem Start von Windows startet.

116 Ich habe schon einen neuen Router gekauft weil ich eine auch ein neues Laptops habe.

117 Ich habe sogar einen neue Router gekauftt, weil das Problem durch den Kauf eines neuen Laptop schlimmer wurde.

118 Ich habe versucht, die verschiedenen Backup-Sätze (Backup alle 2 Stunden), mit und ohne die Streichung der Zielaufwerk.

119 Ich habe versucht, mit dem Startmenü zu deinstallieren und danach auch das $Norton_{Removal_{T}}ool$.

120 Ich habe versucht, mit verschiedenen Backup-Sätze (Backup alle 2 Stunden), mit und ohne Löschung der Ziellaufwerk.

121 Ich habe Vorteil frei upgraden auf NIS 2012.

122 Ich habe Vorteil frei upgraden auf NIS 2012.

123 Ich hatte diese Website für einen kurzen Zeitraum ruhen, bis ich die ID Safe wouldn 't ausfüllen zu unterzeichnen.

124 Ich hatte noch ein Problem mit sowohl PCs:

125 Ich hoffe wirklich, jemand kann mir einige gute Ratschläge geben.

126   Ich brauche mein Computer dringend.
127   Ich hoffe, dass Informationen hilft ihnen weiter
128   Ich höre den Ton "Ausführen", aber nichts wird geöffnet.
129   Ich interessiert, ist das erste Laufwerk C: In dieser Liste enthalten.
130   Ich machte ein Upgrade zu Version 13.
131   Ich möchte das Programm unter http://adguard.com/en/welcome.html ad-guardinstall.exe herunterladen, aber NIS verhindert dies und meldet, dies sei eine Bedrohung.
132   Ich möchte das Programm unter http://adguard.com/en/welcome.html versuchen, aber wenn ich adguardinstall.exe herunterlade, NIS wird es sofort löschen, sagen, dass es eine Bedrohung ist.
133   Ich möchte die andere wieder zu haben.
134   Ich möchte eine andere haben.
135   Ich möchte meine bestehenden Einstellungen zu behalten.
136   Ich möchte nur eben netflix beobachten ohne unglaubliche Ladezeiten.
137   Ich möchte wissen, ob es eine Lösung für gibt , oder es vielleicht bei Norton inoch nicht bekannt ist.
138   Ich muss also über den Datenträger installieren und schauen ob mir jemand dabei helfen kann.
139   Ich muss über Datenträger zu installieren, so ich muss jemand finden, wer kann mir damit helfen.
140   Ich sehe auch System reserviert (D:\) mit Fragezeichen und (E:\) mit Fragezeichen.
141   Ich versuchte verschiedene Backups, mit 2 Stunden Unterschied. Mit oder ohne target drive.
142   Ich verwende diese Informationen, um alles in meine IE9 Browser-Datenschutz-Websites zu blockieren.
143   Ich verwende diesen Informationen blockiert alles, was in meinem IE9 Browser-Datenschutz Websites.
144   Ich verwende diesen Informationen, um in meinem IE9 Browser Privatsphäre zu blockieren.
145   Ich verwende Ghost, weil das immer großartig und einfach funktioniert.
146   Ich werde für eine Woche nicht die Zeit haben, um dies zu tun.
147   Ich müsste von Disk installieren und merkte jemand muss mir dabei helfen.
148   IE9 und NIS Symbolleiste Frage
149   In diesem Fall schlage ich die Registrierung auf einer der Malware entfernen Seiten vor.
150   In diesem Fall schlage ich die Registrierung auf einer Site vor, die Malware entfernt.

151 In diesem Fall schlage ich vor der Registrierung auf einem der Malware entfernen.

152 In fast allen Fällen verursacht werden diese Ereignisse von legitimen Programmen oder Windows Prozesse verursacht und nichts ist zu befürchten.

153 In Zwischenablage kopieren fehlerhaft für Bildschirm "Red Threats erkannt"

154 Ist ein physisches oder elektrisches Trauma aufgetreten?

155 Ist es nur dass es nicht aktualisiert wird (wie NOD32), oder hört es völlig auf zu funktionieren?

156 Ist es nur nicht möglich zu aktualisieren (wie NOD32), oder wird es völlig aufhören zu funktionieren?

157 Ja.

158 Jede Ressource sagt, dass diese Infektion nur auf PCs zutrifft und bietet Lösungen auf PCs, um es zu lösen.

159 Jede Ressource sagt, dass diese Infektion nur für PCs ist und bietet Lösungen auf PCs, um sie zu lösen.

160 Jedesmal wenn ich dieses Programm http: / / adguard.com / en / welcome.html adguardinstall.exe lade, löscht NIS dieses Programm.

161 Jetzt mit der neuen Version nach Re- Abonnements, die nicht NIS 2013 anzeigt sondern nur (Norton Internet Security) ver. 20.2.1.22.

162 Jetzt verwende ich NIS 2013 (Norton Internet Security) in der Version 20.2.1.22.

163 Kann ein Norton Antiviruslizenz übertragen werden von einem Computer auf anderen?

164 Kann ich die Norton Antivirus License Be auf einen anderen Computer übertragen ?

165 Klicken Sie auf das Symbol "Papierkorb kann, dann können Sie auf Ihrem Laptop zu installieren.

166 Könnte es damit zu tun haben, das der automatische Ausgeführungsfunktion ausgeschaltet ist?

167 Kopieren in die Zwischenablage für rot erkannte Bedrohungen auf dem Bildschirm

168 Lassen Sie das Laufwerk C und das Ziellaufwerk auf die Art und Weise, wie sie sind.

169 Live-Chat ist normalerweise die schnellste.

170 Mein Dank...

171 Meine 2. Frage ist:

172 Mit Anzeigen eines Bildes / Image in einer Spam-Nachricht wird es ermöglichen, dass Spammer sie benutzen können und so mit aktiv ist - und viel Erfolg verspricht im versenden von Spam -

173 Mit ctrl Alt löschen und dann bei dem Aufgaben manager auf Start klicken und somit die Prozesse zu beenden.

174 Mit dem Ereignis ist die Informationen: NOF beigefügt.

175 Mit dem Ereignis ist die folgenden Informationen verbunden: Nof.

176 Mit Rechts Klicke wird es geöffnet und zeigt mir ein Menü an um die Installation zu beginnen.

177 Möglicherweise können Sie diese Beschreibung mit dem Kennzeichen /AUXSOURCE= abrufen.

178 Nach der Neuinstallation funktioniert alles wie zuvor, keine spezifischen Tracking Cookie-Informationen in der Kopierfunktion.

179 Neue Ideen?

180 NIS 2012 Datenträger konnte nicht automatisch zu starten werden

181 Norton Ghost IMMER sagt "Gefährdet"

182 Norton macht es einfach sicher, dass nichts, nicht einmal Windows, Probleme durch Unterbrechung von Norton-Operationen verursachen könnte.

183 Norton wird automatisch neue Regeln für das Programm konfigurieren.

184 Norton-Symbolleiste in Google Chrom

185 Nur über alle E-Mails mit Images in der HTML-Code angezeigt werden blockiert, und ich muss manuell "OK" jedes Mal, sie zu sehen.

186 Nur um sie zu informieren, wann Sie das Upgrade auf NMS von Android Market kaufen, wird es nicht in Ihrem Norton Account Produkt angezeigt.

187 Oder drittens gibt es eine Möglichkeit, ich kann einfach nur den neuen Produktschlüssel eingeben und die vorhandene Installation mit einen längeren Zeitraum zu halten?

188 Oder drittens, gibt es eine Möglichkeit dass ich einfach neuen Produktschlüssel eingebe und die vorhandene Installation mit einen längeren Zeitraum habe?

189 Öffnen Sie den Ghost und klicken Sie auf die große Schaltfläche "Erweitert".

190 Öffnen Sie den Ghost und klicken Sie auf die Schaltfläche "Erweitert".

191 Öffnen Sie die Programmsteuerung Fenster in Norton ("Einstellungen" > "Netzwerk" > "Intelligente Firewall" > "Programmregeln").

192 Öffnen Sie diese neu in der Textdatei gespeichert.

193 Programm: C:\Program Files\Norton Online\Engine\2.3.0.7\ccSvcHst.exe R6025 -pure virtual function call.

194 Programm: C:\Program Files\Norton Online\Engine\2.3.0.7\ccSvcHst.exe R6025 -pure virtual function call.

195 Schließlich ich sehe die [Mein Book 2x2TB] (F:) aufgeführt mit einem X in einem roten hexagon.

196 Sie können die Norton Prozesse beenden.

197 Sie können Sie erreichen unter http: / / www.norton.com / Chat.

198 Sie möchten Berichte für das Laufwerk C und das Ziellaufwerk sollte bereits deaktiviert sein und als "Backup-Ziel" aufgelistet sein.

199 Sie möchten Berichte über Laufwerk C und das Ziellaufwerk sollte schon ausgeschaltt und als "Backup-Ziel" aufgelistet sein.

200 Sie sind sehr gut Diese zu lösen.

201 Sie war gehackt worden.

202 Sie wollen auf dem Laufwerk C Berichterstattung und das Ziellaufwerk sollte bereits deaktiviert werden und aufgelistet als "Backup-Ziel".

203 Sie wurde gehackt.

204 Situation: Ein seltsames Verhalten von Windows 7 Pro 64 Bit. Das Programm startet nicht, ohne Netzwerkverbindung.

205 Nachdem ich versuchte mein Handy zu entsperren.

206 So führen Sie die Start der Installation manuell, anstatt Autorun immer aktivieren und deaktivieren.

207 So habe ich jetzt beschränkt sich auf Websites und stattdessen auf Lesezeichen (IE nennt sie Favoriten).

208 So habe ich jetzt mich bei Websites eingeschränkt und setze stattdessen Lesezeichen (IE nennt sie Favoriten).

209 So kann ich die neue Kopie über die vorhandene Version wiederinstallieren?

210 Sobald das getan ist, starten Sie den Computer neu in die SRD, und versuchen Sie, die Sie wiederherzustellen.

211 Sobald es eine Seite geöffnet, ich habe bemerkt, dass die E-Mail nicht sent by a friend...

212 Sobald getan, starten Sie den Computer wieder in die SRD, und versuchen Sie, die Sie wiederherstellen möchten.

213 Sobald sich eine Seite geöffnet hatte, habe ich bemerkt, dass die E-Mail nicht sent by a friend...

214 Sowohl PCs haben Windows 7 mit Norton 360.

215 Starten Sie dann neu und schließen Sie Download und Ausführung des Norton Removal Tools ab, bevor Sie Ihr neues Produkt installieren.

216 Stellen Sie sicher, DASS SIE DIE KORREKTE LAUFWERK LÖSCHEN UND DASS SIE NICHT DEN LAUFWERK LÖSCHEN MIT DER IMAGE DATEI!!!

217 Stellen Sie unbedingt sicher, DASS SIE DIE KORREKTE PLATTE LOESCHEN UND NICHT DIE DISK MIT DER IMAGE-DATEI!!!

218 Trojaner in einen Computer

219 Tun Sie das, es gehört zu diesem Programm.

220 Über den erweiterten Bildschirm sehe ich mein Laufwerk C: Mit einem grünen Kreis und Prüfen von C: Laufwerke mit Fragen markiert.

221 Über den Status Bildschirm sehe ich eine Tabelle mit dem Laufwerk C: Und BU Jobs nach Tag.

222 Über die erweiterten Bildschirm sehe ich mein Laufwerk C: Mit einem grünen Kreis und aktivieren Sie 2 anderen C: Laufwerke mit Fragen.

223 Überprüfen Sie den Link unten, wo die Sicherheitslösungen angezeigt wird, wenn das jetzige Programm nicht kompatibel with. http://www.symantec.com/de/de/Norton/support/kb/web_view.jsp? WV_type=public_web&ssfromlink=true&a

224 Um das Problem zu arbeiten, müssen Sie einen Quick Scan ausführen.

225 Um den kurzen Benutzername festzustellen:

226   Um diese Beschreibung zu verwenden, können Sie die / AUXSOURCE = Flagge verwenden.

227   Und der Link www.norton.com / nms wird verwendet, um die Datei APK herunterzuladen.

228   Und sie müssen den Link www.norton.com / nms Link APK verwenden, um die Datei herunterzuladen.

229   Versuchen Sie folgenses um zu zu sehen, ob es das Problem behebt Einfrieren / hängen bleiben, bis eine dauerhaftere Lösung verfügbar ist:

230   Verwenden Sie diesen Namen, wenn File Guard nachfür Ihrem Administratorkennwort fragt.

231   Vollständiger Systemscan Fehler

232   Vor kurzem wurde mein ROM beschädigt und sperrte mein Telefon.

233   Vor zwei Tagen, als ich auf einenauf einen Link in einer E-Mail geklickt habe, ich dachte sie war von einem Freund.

234   Wählen Sie den Link "Exportergebnisse", um die Prüfung der unten angezeigten Ergebnisse in einer Textdatei:

235   Während der Wiederherstellung der Festplatte gestartet, löschen Sie die Festplatte wiederherstellen.

236   Während Sie auf die Wiederherstellungplatte gebootet haben, löschen Sie die Festplatte, die Sie wiederherstellen.

237   Während vom Wiederherstellungslaufwerk gebootet wird, löschen Sie die wiederherzustellende Festplatte sauber.

238   Warten Sie, bis Ihr Abonnement in ein oder zwei Tage abläuft, öffnen Sie dann NIS und klicken Sie auf Verlängern.

239   Was ist der beste Weg um NIS 2012 zu installieren. über eine exentierende Version?

240   Was ist Norton heißen?

241   Was muss ich tun, um den Status "Gefährdet" verschwinden zu lassen, ohne zu riskieren, meine Backups löschen?

242   Was muss ich tun, um die "Gefährdet" Bedingung verschwinden, ohne zu riskieren, meine Backups löschen?

243   Wechseln Sie hier, und melden Sie sich mit Ihrem Norton Account https: / / manage.norton.com / Account.

244   Wenn der PC bootet zum wiederherstellen, löscht das Programm die HDD und stellt diese wieder her.

245   Wenn er Norton erwirbt, wird es entfernen dann diesen Trojaner Viren von seimem Computer entfernen?

246   Wenn ich auf das Laufwerk klicke, will es öffnen, um die Installation zu starten.

247   Wenn ich ctrl ALT löschen drücke und dann auf "Start Task Manager" klicke und dann zu "Prozessen" gehe.

248   Wenn ich die Flash-Laufwerk in den Computer gestern einfügte für ein Backup auszuführen, die Flash-Laufwerk war leer (keine Dateien).

249   Wenn ich dies mit einem tiefen Scan ausführe, das Schrauben meine CanoScan Toolbox 5.0 Anwendung für meine CanoScan 4400F.

250   Wenn ich eine neue oder andere Registerkarte, die bereits geöffnet ist, und wieder zur ursprünglichen Seite zurückkehre ist diese vollständig geladen und funktioniert gut.

251   Wenn ich eine Webseite öffne (Norton Forum) zeigt mir die NT und IS kein Symbol.

252   Wenn ich Norton oeffne und vollständige Systemprüfung auswaele, zeigt das Scan-Fenster an (in rot):

253   Wenn ich Norton öffne und die Option, führen Sie eine vollständige Systemprüfung durch, wird das Scan Fenster (in rot) angezeigt:

254   Wenn ich Norton öffne und wähle die Option eine vollständige Systemprüfung zu führen, das Scan Fenster zeigt (in rot):

255   Wenn ich Norton starte für ein Komplettscan, wird das Menü in Rot angezeigt.

256   Wenn ich STRG ALT Entf durchführe und klicke auf Task-Manager starten und dann unter Prozesse gehe.

257   Wenn ich Strg, ALT und Entf gleichzeitig drücke, startet der Taskmanager und Prozesse fahren herunter.

258   Wenn Sie auf # 11, machen Sie sich keine Sorgen über die Schaffung einer neuen Partition als Ghost wiederhergestellt werden.

259   Wenn Sie auf # 11, machen Sie sich keine Sorgen über die Schaffung einer neuen Partition weil die Ghost Wiederherstellung das machen wird.

260   Wenn Sie das getan haben, das gehört zu diesem Programm.

261   Wenn Sie das machen, gehört zu diesem Programm.

262   Wenn Sie das zweite Programm downloaden und installieren, würde der minimale Schutz verringert.

263   Wenn Sie eine Einzelhandelsversion von NMS es auch nicht registrieren Sie sich als Norton Produkt in Ihrem Norton Account.

264   Wenn Sie nicht die Erneuerung durchführen und werden ein anderes Produkt verwenden, deinstallieren Sie NAV über die Systemsteuerung in Windows.

265   Wenn Sie nicht verlängern und ein anderes Produkt nutzen werden, sollten Sie NAV von der Windows-Systemsteuerung deinstallieren lassen.

266   Wenn Sie die zweite Programm herunterladen und installieren, konnte das Ergebnis den Schutz verringern.

267   wie kann ich Norton in der Liste finden?

268   Wie kann man am besten eine neue Kopie von NIS 2012 installieren mit einer vorhandene aber zeitlich begrenzten Installation?

269   Wie Sie am besten eine neue Version von NIS 2012 über eine vorhandene, aber zeitlich begrenzte Version installieren ?

270   Wiederherstellen Scheitern von Ghost 15

271   Wiederherstellen mit Ghost 15 gescheitert

272   Wiederherstellung mit einem USB Stick

273   Willkommen, ob sie wiederhergestellt werden können hängt davon ab wie sie fehlen.

274   Willkommen, wie sie wiederhergestellt werden können, wird davon abhängen, wie sie vermisst gegangen.

275   Wo kann ich mein Produkt aktivieren?

276   Wo soll ich meinen Aktivierungsschlüssel holen, oder muss ich oder Google plagen, um meine Appliste zu korrigieren?

Appendix M

Correlations between Error Annotation and Errors Corrected

In a second approach to testing the hypothesis, Spearman's rho was calculated for the importance of each category and errors corrected (sub-categories). This is displayed in Table M.1 with column A as categories of actual post-editing behaviour to compared to column B self-reported importance of category before the experiment and C as self-reported importance of category after the experiment.

A	B	C
category	**content (before)**	**content (after)**
missing info	0.46	0.12
extra info	0.07	0.34
untranslated	-0.10	0.43
	grammar (before)	**grammar (after)**
adjective	-0.46	-0.14
capitalisation	-0.60	0.14
conjunction	0.18	0.12
determiner	-0.60	0.32
gender	0.16	0.03
noun	-0.02	0.17
number	-0.17	0.14
phrasal ordering	0.10	0.40
preposition	-0.02	-0.04
pronoun	0.81	0.82
verb	0.06	0.36
	spelling (before)	**spelling (after)**
spelling	0.31	0.17
	punctuation (before)	**punctuation (after)**
punctuation	0.15	0.24
format	0.22	-0.31

TABLE M.1: Spearman's rho of self-reported error tolerances and actual post-editing behaviour